



The  
Patent  
Office



INVESTOR IN PEOPLE

EP 99/10297

REC'D 21 MAR 2000 The Patent Office  
Concept House  
WIPO PCT Cardiff Road  
Newport  
South Wales  
NP10 8QQ

EPO - DG 1

14. 02. 2000

(59)

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

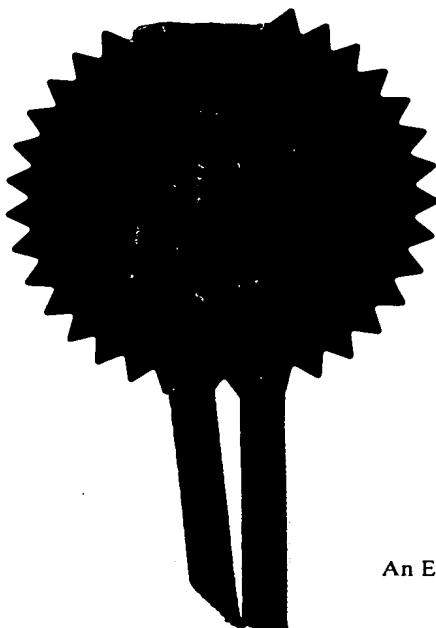
In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

PRIORITY DOCUMENT

Signed *Andrew Gersey*

Dated 11 November 1999



**THIS PAGE BLANK (USPTO)**

21 DEC 1998

For Official use

9828217.1

22DEC98 E413657-1 D02029  
P01/7700 0.00 - 9828217.1

Your Reference: B45168

**Notes**

Please type, or write in dark ink using CAPITAL letters. A prescribed fee is payable for a request for grant of a patent. For details, please contact the Patent Office (telephone 071-483 4700).

Rule 16 of the Patents Rules 1990 is the main rule governing the completion and filing of this form.

- ② Do not give trading styles, for example, 'Trading as XYZ company', nationality or former names, for example, 'formerly (known as) ABC Ltd' as these are not required.

**Warning**

After an application for a Patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977 and will inform the applicant if such prohibition or restriction is necessary. Applicants resident in the United Kingdom are also reminded that under Section 23, applications may not be filed abroad without written permission unless an application has been filed not less than 6 weeks previously in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction revoked.

# The Patent Office

## Request for grant of a Patent

### Form 1/77 Patents Act 1977

**① Title of invention**

1 Please give the title of the invention **VACCINE**

**② Applicant's details**  
☐ First or only applicant

2a If you are applying as a corporate body please give:  
 Corporate Name **Université Libre de Bruxelles**

Country (and State of incorporation, if appropriate) **Belgium**

2b If you are applying as an individual or one of a partnership please give in full:

Surname

Forenames

2c In all cases, please give the following details:

Address: **Faculté des Sciences, Rue de l'Industrie 24, B - 1400 Nivelles, Belgium**

UK postcode (if applicable)

Country **Belgium**  
 ADP number **5800974002**  
 (if known)

05663620002

2d, and 2f: If there are further applicants please provide details on a separate sheet of paper

☐ **Second applicant (if any)**  
2d If you are applying as a corporate body please give:  
Corporate Name

Country (and State of  
Incorporation, if appropriate)

2e If you are applying as an individual or one of a partnership please  
give in full:

Surname:

Forenames:

2f In all cases, please give the following details:

Address:

UK postcode

(if applicable)

Country

ADP number

(if known)

3 An address for  
service in the United Kingdom  
must be supplied

Please mark correct box

3 **Address for service details**

3a Have you appointed an agent to deal with your application?

Yes ☒

No ☐ go to 3b



please give details below

Agent's name **MARCUS J W DALTON**

Agent's address **SmithKline Beecham  
Corporate Intellectual Property  
New Horizons Court  
Great West Road  
Brentford, Middlesex**

Postcode **TW8 9EP**

Agent's ADP  
number

**04122313006**

3b: If you have appointed an agent,  
all correspondence concerning  
your application will be sent to the  
agent's United Kingdom address.

3b If you have not appointed an agent, please give a name and address in  
the United Kingdom to which all correspondence will be sent:

Name

Address

Postcode  
number

ADP number  
(if known)

Daytime telephone  
(if available)

**Please mark correct box**

*Please give the date in all number format, for example, 31/05/90 for 31 May 1990*

<b>④ Reference number</b>		
4. Agent's or applicant's reference number <i>(if applicable)</i>		
<b>⑤ Claiming an earlier application date</b>		
5. Are you claiming that this application be treated as having been filed on the date of filing of an earlier application?		
<div style="display: flex; justify-content: space-between;"> <span>Yes <input type="checkbox"/></span> <span>No <input checked="" type="checkbox"/>  go to 6</span> </div>		
<div style="text-align: center;"> <p><i>please give details below</i></p> </div>		
<input type="checkbox"/> number of earlier application or patent number		
<input type="checkbox"/> filing date <span style="margin-left: 20px;"><i>(day)</i></span> <span style="margin-left: 20px;"><i>month</i></span> <span style="margin-left: 20px;"><i>year</i></span>		
<input type="checkbox"/> and the Section of the Patents Act 1977 under which you are claiming:		
<div style="display: flex; justify-content: space-between;"> <span>15(4) (Divisional) <input type="checkbox"/></span> <span>8(3) <input type="checkbox"/></span> <span>12(6) <input type="checkbox"/></span> <span>37(4) <input type="checkbox"/></span> </div>		
<b>⑥ Declaration of priority</b>		
6. If you are declaring priority from previous application(s), please give:		
Country of Filing	Priority application number <i>(if known)</i>	Filing Date <i>(day, month, year)</i>

- any applicant is not an inventor
- there is an inventor who is not an applicant, or
- any applicant is a corporate body.

8 Please supply duplicates of claim(s), abstract, description and drawings).

Please mark correct box(es).

9 You or your appointed agent (see Rule 90 of the Patents Rules 1990) must sign this request.

Please sign here *a*

A completed fee sheet should preferably accompany the fee.

## 7 Inventorship

7. Are you (the applicant or applicants) the sole inventor or the joint inventors?

Please mark correct box

Yes ☐ No ☒ A Statement of Inventorship on Patents form 7/77 will need to be filed (see Rule 15).

## 8 Checklist

8a Please fill in the number of sheets for each of the following types of document contained in this application.

Continuation sheets for this Patents Form 1/77

Claim(s) 3

Description 31

Abstract

Drawing(s) 27

8b Which of the following documents also accompanies the application?

Priority documents (please state how many)

Translation(s) of Priority documents (please state how many)

Patents Form 7/77 - Statement of Inventorship and Right to Grant

Patents Form 9/77 - Preliminary Examination Report

Patents Form 10/77 - Request for Substantive Examination

## 9 Request

I/We request the grant of a patent on the basis of this application.

Signed

*Marcus J. Dalton*  
MARCUS J W DALTON  
Chartered Patent Attorney  
Attorney for the Applicant

Date: 21/12/98

(day month year)

Please return the completed form, attachments and duplicates where requested, together with the prescribed fee to either;

☐ The Comptroller or  
The Patent Office  
Cardiff Road  
Newport  
Gwent  
NP9 1RH

☒ The Comptroller  
The Patent Office  
25 Southampton Buildings  
London  
WC2A 1AY

## VACCINE

### FIELD OF INVENTION

This invention relates to a general method for detecting pathogenic strains of  
5 bacteria which harbour a type III secretion system, and characterising regions of the  
chromosome of said strain where virulence genes reside. More particularly, this  
invention relates to the method as applied to the pathogen *Bordetella pertussis*.  
Furthermore, the invention relates to newly identified polynucleotides within these  
regions, virulent polypeptides encoded by them and to the use of such polynucleotides  
10 and polypeptides, and to their production. More particularly the polynucleotides and  
polypeptides of the present invention relate to the BopN, Orf1, Orf2, Orf3, Orf4, Orf5,  
Orf6, Orf7, Orf8, Orf9, Orf10, Orf11, Orf12, Orf13, Orf14, and Orf15 effector proteins  
of *Bordetella pertussis*.

### 15 BACKGROUND OF THE INVENTION

#### Type III secretion systems:

Pathogenic bacteria invade many different niches in a broad host range and cause  
a wide variety of syndromes. It is due to this fact that it was believed previously that each  
20 disease might be induced by a distinct molecular mechanism. However, the spectrum of  
such mechanisms is not as broad as first imagined; rather, bacteria exploit a number of  
common molecular tools to achieve a range of goals. Among these tools are type III  
secretion systems, which provide a means for bacteria to target virulence factors directly  
at host cells. These factors then tamper with host cell functions to the pathogens' benefit.

25

The type III export system is responsible for secretion of *Salmonella* and *Shigella*  
invasion and virulence factors, Enteropathogenic *Escherichia coli* (EPEC) signal  
transduction molecules, virulence factors in several plant pathogens (for instance  
*Xanthomonas campestris* pv. *vesicatoria* [Fenselau et al., 1992]) and Yops proteins in  
30 *Yersinia*. Yops export mechanism has been the most intensively investigated type III

secretion apparatus (see for instance: Allaoui et al., 1994; Bergman et al., 1994). In this system, more than 20 different Ysc/Lcr proteins, all encoded by the virulence plasmid pYV, are presumed to compose a secretion-channel-spanning the *Yersinia* cell envelope. Besides these elements involved in the secretion machinery, the pYV plasmid codes for the Yops proteins which are the secreted substrates and appear as the actual effectors of virulence.

Comparative studies of type III secretion systems originating from different species reveal that the components of the secretion machinery are strikingly conserved (Gygi et al., 1995; Bogdanove et al., 1996). In addition, homologs have been found in determinants which take part in flagellar assembly, indicating that this secretion pathway may be involved in surface organelle biosynthesis (Ramakrishnan et al., 1991).

In contrast, however, the secreted substrates share no similarities, except in few cases. Therefore, the abandoned concept of a distinct molecular mechanism corresponding to each disease could reappear at the level of effector proteins.

### Pathogenicity island

Pathogenicity islands have emerged as a novel theme in the field of bacterial virulence. Although they can comprise type III secretion systems they do not exclusively do so.

Early in the search for virulence genes, it was observed that many of these genes resided on plasmids. However, numerous virulence genes were also found on the chromosome. Surprisingly, similarly as when they are found on plasmids, the chromosomal virulence genes often clustered in functionally related groups also. Such groups of virulence genes gave rise to the concept of pathogenicity islands (Pais) which can be defined as compact, distinct genetic units carrying virulence genes. These units,



often flanked by direct repeats, occupy large chromosomal regions (often > 30 kb) and are present in pathogenic strains, whilst being absent or sporadically distributed in less-pathogenic (or non-pathogenic) strains of a bacterial species. These DNA segments are frequently associated with tRNA genes and/or insertion sequence (IS) elements at their boundaries. In addition, their G+C content often differs from that of host bacterial DNA, suggesting a foreign origin.

Pathogenicity islands have been discovered in an increasing number of bacterial pathogens, including different categories of *E. coli*, *Salmonella typhimurium*, *Yersinia* spp, *Helicobacter pylori*, *Vibrio cholera* etc.

The first intensively studied pathogenicity islands were Pai I and Pai II, which encode the haemolysin determinants of uropathogenic *E. coli*. These two Pairs, are flanked by direct repeats and can be deleted from the chromosome at frequencies of  $10^{-4}$ , resulting in non-virulent mutant strains. Another pathogenicity island of 35 kb has recently been identified on the chromosome of enteropathogenic *E. coli* (EPEC) and was found to encode all known determinants involved in the so-called "attaching and effacing" (AE) lesion formation. This region was therefore referred to as "locus of enterocyte effacing" (LEE). Despite the fact that uropathogenic and enteropathogenic *E. coli* cause completely different infectious diseases, Pai I of the uropathogenic strains and the LEE locus of EPEC are inserted at exactly the same positions into the *E. coli* chromosome.

While some authors support a definition of pathogenicity islands which necessarily includes its chromosomal location, others have extended the concept to blocks of virulence genes, regardless of their location in chromosomes, plasmids or phages. The fact that, on one hand, phages and plasmids can easily insert into and excise from the chromosome and, on the other, that cryptic origins of plasmid replication, or phage related sequences were detected in Pairs, prompted us to adopt the latter and less restrictive definition.

The pathogenicity islands (PAIs) which code for a type III secretion system encompass genes that divide into two classes, I and II. Class I encompasses the genes coding for the secretion machinery components and their regulators of expression, class II encompasses the genes encoding secreted effector proteins. Both *Yersinia lcrD* and *yscU* belong to class I. The precise functions of class I determinants is not well understood. Although it is sometimes not straightforward to make a clear distinction between class I and class II components, genes of class I can be identified as being present in many different species, and a comparison of their respective gene sequences indicate that equivalent genes share a significant (*yscI*, *yscO*) or even high level (*lcrD*, *yscU*, *yscN*) of sequence similarity (Hueck, 1998).

The second class of genes (class II) codes for proteins which constitute the substrate secreted by the translocon. These proteins appear as the actual effectors of virulence and are referred to as target proteins, virulence effector proteins or, simply, effectors. In contrast to the situation prevailing in class I gene products, the effectors share no, or very weak, similarities between species. Effector proteins are those which present the best biological, vaccine and diagnostic potentialities.

The inventors have realised that the clustering of class I and class II genes inside a single pathogenicity island, offers the opportunity of conveniently finding and characterising unknown class II genes by targeting class I genes which can be identified using a known sequence of one of their numerous orthologues.

### ***Bordetella pertussis***

Whooping cough is a disease caused by infection by *Bordetella pertussis*, and is a serious and debilitating human disease particularly in young children. Although whole cell and acellular vaccines are available that are effective against the disease, there

remains a need for the identification of further highly purified pertussis proteins that could be used in a more efficacious pertussis vaccine.

Although many pertussis virulence associated factors are known such as pertussis toxin, filamentous haemagglutinin, pertactin, which have been included in various acellular vaccines, there is no convenient genetic method for identifying further virulence factors using the pertussis genome (short of laboriously sequencing the whole genome). Although class I type III secretion system virulence genes have recently been shown to exist in *B. bronchiseptica* and *B. pertussis* (Yuk *et al.*, 1998), there has been no complete analysis of a pathogenicity island in *Bordetella*, and the identity and characterisation of effector genes within such a pathogenicity island has been unknown up until the present invention.

## SUMMARY OF THE INVENTION

In one aspect, the invention relates to a method for the identification of new virulence genes in bacterial strains containing a type III secretion system. In particular, the invention allows the identification of the effector virulence genes associated within a pathogenicity island containing the genes for the type III secretion system. Another aspect of the invention a method for the identification of pathogenic bacterial strains containing a type III secretion system. Another aspect of the invention relates to *Bordetella pertussis* BopN, Orf1, Orf2, Orf3, Orf4, Orf5, Orf6, Orf7, Orf8, Orf9, Orf10, Orf11, Orf12, Orf13, Orf14, Orf15 effector proteins, and the respective polynucleotide sequences encoding them.

Although the general concepts of type III secretion systems and pathogenicity islands have been reported, the problem of how simply and reliably to identify whether any given organism has such cell machinery has not been accomplished until now. Such a method is extremely useful to establish whether a given strain has a type III secretion

system within a pathogenicity island, to characterise unknown virulence genes within the pathogenicity island, and to use in quick diagnostic methods for determining whether a cultured bacterial strain containing a type III secretion system is pathogenic.

5 In the present invention, a novel, general method is described to achieve the above aims. More specifically, the invention utilises a method that employs ideally-suited primers designed specifically from the sequence of the virulent *Yersinia enterocolitica lcrD* gene as a target sequence. The presence of a type III secretion system within a pathogenicity island in *Bordetella pertussis* was discovered, and every gene  
10 within the pathogenicity island was characterised.

#### BRIEF DESCRIPTION OF THE DRAWINGS

15 **Fig. 1.** Nucleotide and deduced amino acid sequences of the cloned 152 bp amplicon. The primers involved in the original amplification, the subsequent nested PCR, and the gene library screening are all derived from this sequence, and listed specifically in Table 1.

20 **Fig. 2.** PileUp figure from the deduced amino acid sequences homologous to *Yersinia LcrD*. Abbreviations: BbuFlhA = *Borrelia burgdorferi* FlhA; TpaFlhA = *Treponema pallidum* FlhA; BsuFlhA = *Bacillus subtilis* FlhA; CjeFlbA = *Campylobacter jejuni* FlbA; HpyFlhA = *Helicobacter pylori* FlhA; EcoFlhA = *Escherichia coli* FlhA; StyFlhA = *Salmonella typhimurium* FlhA; YenFlhA = *Yersinia enterocolitica* FlhA;  
25 PmiFlhA = *Proteus mirabilis* FlhA; CcrFlbF = *Caulobacter crescentus* FlbF; EcoFhiA = *Escherichia coli* FhiA; EamHrpI = *Erwinia amylovora* HrpI; PsyHrpI = *Pseudomonas syringae* HrpI; ECEPSepA = Enteropathogenic *Escherichia coli* SepA; StySsaV = *Salmonella typhimurium* SsaV; RsoHrpO = *Ralstonia solanacearum* HrpO; XcaHrpC2 = *Xanthomonas campestris* HrpC2; SflMxiA = *Shigella flexneri* MxiA; StyInvA =  
30 *Salmonella typhimurium* InvA; PaePcrD = *Pseudomonas aeruginosa* PcrD; YenLcrD =

*Yersinia enterocolitica* LcrD; BpeBcrD = *Bordetella pertussis* BcrD; CpsTtsB = *Chlamydia psittaci* TtsB.

**Fig. 3.** Organization of the *Bordetella pertussis* pathogenicity island (Pai). Four house keeping genes (hatched boxes) and the transposase gene of IS481 (black box) are surrounding the Pai. The Pai consists of genes coding for determinants involved in the secretory apparatus and its regulation (class I genes, in grey boxes) as well as ORFs which putitively code for effector proteins (class II genes, in white boxes). Letters indicate the respective class I *bsc* genes whereas numbers correspond to the class II ORFs listed in Table 3.

**Fig. 4.** PileUp figure from the deduced amino acid sequences homologous to *Yersinia* YscU. Abbreviations: BbuFlhB = *Borrelia burgdorferi* FlhB; TpaFlhB = *Treponema pallidum* FlhB; EcoFlhB = *Escherichia coli* FlhB; StyFlhB = *Salmonella typhimurium* FlhB; PmiFlhBpart = partial *Proteus mirabilis* FlhB; YenFlhB = *Yersinia enterocolitica* FlhB; BsuFlhB = *Bacillus subtilis* FlhB; HpyFlhB = *Helicobacter pylori* FlhB; AtuFlhB = *Agrobacterium tumefaciens* FlhB; CcrPodW = *Caulobacter crescentus* PodW; SflSpa40 = *Shigella flexneri* Spa40; StySpaS = *Salmonella typhimurium* SpaS; EcoEscU = *Escherichia coli* EscU; StySsaU = *Salmonella typhimurium* SsaU; BpeBscU = *Bordetella pertussis* BscU; YenYscU = *Yersinia enterocolitica* YscU; RsoHrpN = *Ralstonia solanacearum* HrpN; XcaOrf0part = partial *Xanthomonas campestris* Orf0; EamHrcU = *Erwinia amylovora* HrcU; EheHrcUpart = partial *Erwinia herbicola* HrcU; PsyHrpY = *Pseudomonas syringae* HrpY; CpsOrf1 = *Chlamydia psittaci* Orf1.

**Fig. 5.** The DNA sequence of the *Bordetella pertussis* genome comprising the type III secretion system pathogenicity island. Reference should be made to tables 2, 3, and 4 and Fig. 3 for information regarding open reading frames.

## DESCRIPTION OF THE INVENTION

Type III secretion systems identified to date are encoded by either chromosomal or plasmidic pathogenicity island genes. However, no where in the prior art was it realised that the conservation of genes encoding class I components of type III secretion systems and the clustering of these genes with effector protein coding sequences offered the opportunity for detecting unidentified target proteins involved in host colonisation. Such proteins would be potentially valuable in both vaccinal and diagnostic fields.

Although the known sequence of a gene encoding any conserved (class I) type III secretion machinery protein can be used in performing this invention, the *lcrD* gene is preferred. The chosen gene will act as a target for detecting unidentified pathogenicity islands in related bacterial species. The *lcrD* gene from *Yersinia* is preferred as it codes for the archetype of the recently identified LcrD/FliB family of proteins. Members of this family are involved in host cell invasion, virulence in several phytopathogenic bacteria or in flagellar assembly. *lcrD* is preferred because the LcrD protein, and consequently the gene encoding it, is one of the most conserved determinants of the secretion machinery. Additionally, multiple amino acid comparisons have shown that the classification of the LcrD family members can be split into two main subfamilies, which, interestingly, can be correlated with the functions assigned to these proteins of each subfamily. One subfamily encompasses all the motility-involved proteins, while the other encompasses all the virulence-related determinants. This observation is illustrated in Fig. 2 (and mentioned in Gyri *et al.* (1995) & Bogdanove *et al.* (1996)). Thus, if an unknown *lcrD* homologous gene is identified, it may, after being routinely sequenced, be classified as a virulence or a flagellar gene. Once the pathogenicity island is identified, this simple test would therefore define whether the search for other virulence genes on the pathogenicity island should be initiated.

The preferred method for identifying unknown pathogenicity islands comprising a type III secretion system is by:

- 5 i) identifying two highly conserved regions of the target protein sequence (preferably of LcrD). Preferably, both regions should contain conserved amino acids which are encoded by the fewest number of codon possibilities e.g. Methionine (ATG being the only possibility) or Tryptophan (TGG being the only possibility). This minimises the number of permutations in both degenerate primer sets that are designed in the next stage of the process, thus ensuring a greater probability that each primer set will specifically anneal to the unknown *lcrD*-equivalent gene (thereby minimising background non-specific interactions). Most preferably, regions should also be chosen that are clearly distinguishable from the paralogue *flhA* flagellar genes, present in all flagellated bacterial strains.
- 10 ii) designing a degenerate set of primers for both of the chosen regions such that a) the primers are at least 15 bases long, preferably 20-30 bases long, and still more preferably 21-23 bases long, b) they are degenerate at bases that can be more than one type of nucleotide whilst still encoding the same amino acid (due to the degeneracy of codon usage for amino acids), but no more degenerate than is required to cover all permutations for the amino acid region selected, and c) the primer set that encodes the more N-terminal region of the chosen protein should correspond to the coding strand of its corresponding double-stranded DNA sequence, and the set that encodes the more C-terminal region should correspond to the complementary strand of the corresponding double-stranded DNA sequence.
- 15 20 iii) synthesising the degenerate primer sets of step ii) using conventional DNA synthesis methods well known in the art.
- iv) purifying the primer sets of step iii)
- 25 v) adding both the primer sets and a sample containing nucleic acid from a bacterial strain (preferably a cell sample of the bacterial species itself) together in appropriate quantities and in an appropriate buffer in order to perform a polymerase chain reaction (PCR)
- 30 vi) performing a PCR reaction in order to amplify the region of the gene between the two primers (conditions for performing the PCR reaction can be optimised using techniques well known in the art)

vii) observing the reaction products on a gel (preferably an agarose gel) for an amplified product of the size expected; if no such product is present, the bacterial strain is unlikely to use a type III secretion system; if such a product is present, the bacterial strain is likely to have a type III secretion system, and is likely to be pathogenic.

5

The preferred method for confirming that the amplified product actually corresponds to a virulence gene is by carrying out steps i)-vii) above (where the target protein is LcrD) and then:

- 10      viii) optionally separating the product of correct size from any background products of incorrect size by removing the correct band from the gel, purifying the product by conventional means, and amplifying the product once more with the two degenerate primer sets in another PCR reaction (under preferably more stringent PCR conditions) [this step is required should the product of step vii) not be pure enough for direct cloning];
- 15      ix) inserting the DNA fragment by conventional means into a vector which is capable of being sequenced, and sequencing the fragment;
- x) comparing the deduced amino acid sequence of ix) with that of known members of the LcrD/FliB family of proteins to associate the amplified product as being part of either a virulence or a flagellar gene.

20

And optionally:

- xi) using the internal sequence of the fragment to design primers that are the exact sequence of, and specific to, the unknown *lcrD*-equivalent gene.
- xii) using the primers of xi) firstly to screen a genomic library of the organism for
- 25      positive clones;
- xiii) isolating the clones of xii), and sequence one or more of said clones;
- xiv) scanning the sequence of one clone (and overlapping sequences of other clones) to search for an open reading frame which is approximately the same size as *lcrD* (approximately 2100bp), and encodes a protein homologous to LcrD



xv) ascertaining whether the LcrD-equivalent protein is more homologous with the *flbF* (flagellar protein secretion) gene family or the *lcrD* (type III secretion system pathogenicity island) gene family.

5           The preferred method for characterising the whole pathogenicity island and defining unidentified virulence effector genes is by carrying out steps i)-xv) above (where the target protein is LcrD) and then:

xvi) if the sequence is more homologous with the *lcrD* gene family, designing primers at either extreme of the gene sequence already ascertained, and scanning and  
10           sequencing the genomic library (using a standard chromosome walking strategy – where the insert boundaries of an original clone serves as a probe for screening and cloning adjacent regions) to sequence eventually the whole of the pathogenicity island (both boundaries of which will be defined by the presence of either direct or inverted repeats, or insertion sequences, or the presence of house-keeping genes)

15       xvii) defining unidentified virulence effector genes within the sequenced pathogenicity island

xviii) cloning, expressing and characterising the virulence genes of xvii) which encode virulence effector proteins of the organism

20

### Definitions

“Pertussis pathogenicity proteins” refers generally to polypeptides having the amino acid sequence encoded by the genes defined in tables 2 and 3, or an allelic variant thereof. These proteins are: BcrD, BcrH, BscC, BscD, BscE, BscF, BscI, BscJ, BscK,  
25   BscL, BscN, BscO, BscP, BscQ, BscR, BscS, BscT, BscU, BscV, BrpL, BopN, Orf1, Orf2, Orf3, Orf4, Orf5, Orf6, Orf7, Orf8, Orf9, Orf10, Orf11, Orf12, Orf13, Orf14, Orf15.

“Pertussis pathogenicity genes” refers to polynucleotides having the nucleotide  
30   sequence defined in tables 2 and 3, or allelic variants thereof and/or their complements.

These genes are: *bcrD*, *bcrH*, *bscC*, *bscD*, *bscE*, *bscF*, *bscI*, *bscJ*, *bscK*, *bscL*, *bscN*, *bscO*, *bscP*, *bscQ*, *bscR*, *bscS*, *bscT*, *bscU*, *bscV*, *brpL*, *bopN*, *orf1*, *orf2*, *orf3*, *orf4*, *orf5*, *orf6*, *orf7*, *orf8*, *orf9*, *orf10*, *orf11*, *orf12*, *orf13*, *orf14*, *orf15*.

- 5           “Polypeptide” refers to any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres. “Polypeptide” refers to both short chains, commonly referred to as peptides, oligopeptides or oligomers, and to longer chains, generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene-encoded amino acids.
- 10       “Polypeptides” include amino acid sequences modified either by natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-
- 15       chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched and branched cyclic polypeptides may result
- 20       from posttranslational natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond
- 25       formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as
- 30       arginylation, and ubiquitination. See, for instance, PROTEINS - STRUCTURE AND

MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York, 1993 and Wold, F., Posttranslational Protein Modifications: Perspectives and Prospects, pgs. 1-12 in POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, 5 1983; Seifter *et al.*, "Analysis for protein modifications and nonprotein cofactors", *Meth Enzymol* (1990) 182:626-646 and Rattan *et al.*, "Protein Synthesis: Posttranslational Modifications and Aging", *Ann NY Acad Sci* (1992) 663:48-62.

"Polynucleotide" generally refers to any polyribonucleotide or 10 polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. "Polynucleotides" include, without limitation single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid 15 molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, "polynucleotide" refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The term polynucleotide also includes DNAs or RNAs containing one or more modified bases and DNAs or RNAs with backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases 20 such as inosine. A variety of modifications has been made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically or metabolically modified forms of polynucleotides as typically found in nature, as well as the chemical forms of DNA and RNA characteristic of viruses and cells. "Polynucleotide" also embraces relatively short polynucleotides, often referred to as oligonucleotides.

25

"Variant" as the term is used herein, is a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide respectively, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant 30 may or may not alter the amino acid sequence of a polypeptide encoded by the reference

polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions (preferably conservative), additions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques or by direct synthesis. Variants should retain one or more of the biological activities of the reference polypeptide. For instance, they should have similar antigenic or immunogenic activities as the reference polypeptide. Antigenicity can be tested using standard immunoblot experiments, preferably using polyclonal sera against the reference polypeptide. The immunogenicity can best be tested by measuring antibody responses (using polyclonal sera generated against the variant polypeptide) against purified reference polypeptide in a standard ELISA test. Preferably, a variant would retain all of the above biological activities.

20

"Identity" is a measure of the identity of nucleotide sequences or amino acid sequences. In general, the sequences are aligned so that the highest order match is obtained. "Identity" *per se* has an art-recognized meaning and can be calculated using published techniques. See, e.g.: (COMPUTATIONAL MOLECULAR BIOLOGY, Lesk, A.M., ed., Oxford University Press, New York, 1988; BIOCOMPUTING: INFORMATICS AND GENOME PROJECTS, Smith, D.W., ed., Academic Press, New York, 1993; COMPUTER ANALYSIS OF SEQUENCE DATA, PART I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; SEQUENCE ANALYSIS IN MOLECULAR BIOLOGY, von Heijne, G., Academic Press, 1987; and SEQUENCE ANALYSIS PRIMER, Gribskov, M. and Devereux, J., eds., M Stockton Press, New

30

York, 1991). While there exist a number of methods to measure identity between two polynucleotide or polypeptide sequences, the term "identity" is well known to skilled artisans (Carillo, H., and Lipton, D., *SIAM J Applied Math* (1988) 48:1073). Methods commonly employed to determine identity or similarity between two sequences include, but are not limited to, those disclosed in Guide to Huge Computers, Martin J. Bishop, ed., Academic Press, San Diego, 1994, and Carillo, H., and Lipton, D., *SIAM J Applied Math* (1988) 48:1073. Methods to determine identity and similarity are codified in computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, GCG program package (Devereux, J., et al., *Nucleic Acids Research* (1984) 12(1):387), BLASTP, BLASTN, FASTA (Atschul, S.F. et al., *J Molec Biol* (1990) 215:403). Most preferably, the program used to determine identity levels was the GAP program, as was used in the Examples below.

As an illustration, by a polynucleotide having a nucleotide sequence having at least, for example, 95% "identity" to a reference nucleotide sequence is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include on average up to five point mutations per each 100 nucleotides of the reference nucleotide sequence. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. These mutations of the reference sequence may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence.

### Polypeptides of the invention

In one aspect, the present invention relates to *Bordetella* pathogenicity proteins (or polypeptides). The *Bordetella* pathogenicity polypeptides include the polypeptides encoded by the genes defined in tables 2 and 3; as well as polypeptides comprising the amino acid sequence encoded by the genes defined in tables 2 and 3; and polypeptides comprising the amino acid sequence which have at least 75% identity to that encoded by the genes defined in tables 2 and 3 over their entire length, and preferably at least 80% identity, and more preferably at least 90% identity. Those with 95-99% identity are highly preferred.

10        The *Bordetella* pathogenicity polypeptides may be in the form of the "mature" protein or may be a part of a larger protein such as a fusion protein. It may be advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification such as multiple histidine residues, or an additional sequence for stability during recombinant production.

15        Fragments of the *Bordetella* pathogenicity polypeptides are also included in the invention. A fragment is a polypeptide having an amino acid sequence that is the same as part, but not all, of the amino acid sequence of the aforementioned *Bordetella* pathogenicity polypeptides. As with *Bordetella* pathogenicity polypeptides, fragments may be "free-standing," or comprised within a larger polypeptide of which they form a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the invention, include, for example, fragments from about amino acid number 1-20, 21-40, 41-60, 61-80, 81-100, and 101 to the end of *Bordetella* pathogenicity polypeptide. In this context "about" includes the particularly recited ranges larger or  
20        smaller by several, 5, 4, 3, 2 or 1 amino acid at either extreme or at both extremes.

25        Preferred fragments include, for example, truncation polypeptides having the amino acid sequence of *Bordetella* pathogenicity polypeptides, except for deletion of a continuous series of residues that includes the amino terminus, or a continuous series of residues that  
30        includes the carboxyl terminus and/or transmembrane region or deletion of two continuous

series of residues, one including the amino terminus and one including the carboxyl terminus. Also preferred are fragments characterized by structural or functional attributes such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions. Other preferred fragments are biologically active fragments. Biologically active fragments are those that mediate Bordetella pathogenicity protein activity, including those with a similar activity or an improved activity, or with a decreased undesirable activity. Also included are those that are antigenic or immunogenic in an animal, especially in a human.

Preferably, all of these polypeptide fragments retain the biological activity (for instance antigenic or immunogenic) of the Bordetella pathogenicity protein, including antigenic activity. Variants of the defined sequence and fragments also form part of the present invention. Preferred variants are those that vary from the referents by conservative amino acid substitutions -- i.e., those that substitute a residue with another of like characteristics. Typical such substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; and among the basic residues Lys and Arg; or aromatic residues Phe and Tyr. Particularly preferred are variants in which several, 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination. Most preferred variants are naturally occurring allelic variants of Bordetella pathogenicity polypeptide present in strains of *Bordetella pertussis*.

The Bordetella pathogenicity polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

**Polynucleotides of the invention**

Another aspect of the invention relates to Bordetella pathogenicity polynucleotides. Bordetella pathogenicity polynucleotides include isolated polynucleotides which encode the Bordetella pathogenicity polypeptides and fragments respectively, and polynucleotides  
5 closely related thereto. More specifically, Bordetella pathogenicity polynucleotides of the invention include a polynucleotide comprising the nucleotide sequence of genes defined in table 2 or 3, encoding a Bordetella pathogenicity polypeptide. Bordetella pathogenicity polynucleotides further include a polynucleotide comprising a nucleotide sequence that has  
10 at least 75% identity over its entire length to a nucleotide sequence encoding the Bordetella pathogenicity polypeptide encoded by the genes defined in tables 2 and 3, and a polynucleotide comprising a nucleotide sequence that is at least 75% identical to that of the genes defined in tables 2 and 3. In this regard, polynucleotides at least 80% identical are particularly preferred, and those with at least 90% are especially preferred. Furthermore, ~~those with at least 95% are highly preferred and those with at least 98-99%~~  
15 ~~are most highly preferred, with at least 99% being the most preferred.~~ Also included under Bordetella pathogenicity polynucleotides is a nucleotide sequence which has sufficient identity to a nucleotide sequence of a gene defined in tables 2 and 3 to hybridize under conditions ~~useable for amplification or for use as a probe or marker.~~ The invention also provides polynucleotides which are complementary to such Bordetella pathogenicity  
20 polynucleotides.

The nucleotide sequence encoding Bordetella pathogenicity polypeptide encoded by the genes defined in tables 2 and 3 may be identical to the polypeptide encoding sequence contained in the genes defined in tables 2 or 3, or it may be a sequence, which as  
25 a result of the redundancy (degeneracy) of the genetic code, also encodes the polypeptide encoded by the genes defined in tables 2 and 3 respectively.

~~When the polynucleotides of the invention are used for the recombinant~~  
production of Bordetella pathogenicity polypeptide, the polynucleotide may include the  
30 coding sequence for the mature polypeptide or a fragment thereof, by itself; the coding



sequence for the mature polypeptide or fragment in reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-, or pro- or prepro-protein sequence, or other fusion peptide portions. For example, a marker sequence which facilitates purification of the fused polypeptide can be encoded. In certain preferred  
5   embodiments of this aspect of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al.*, *Proc Natl Acad Sci USA* (1989) 86:821-824, or is an HA tag, or is glutathione-s-transferase. The polynucleotide may also contain non-coding 5' and 3' sequences, such as transcribed, non-translated sequences, splicing and polyadenylation signals, ribosome binding sites and  
10   sequences that stabilize mRNA.

Further preferred embodiments are polynucleotides encoding *Bordetella* pathogenicity protein variants comprising the amino acid sequence of the *Bordetella* pathogenicity polypeptide encoded by the genes defined by tables 2 and 3 respectively in  
15   which several, 10-25, 5-10, 1-5, 1-3, 1-2 or 1 amino acid residues are substituted, deleted or added, in any combination. Most preferred variant polynucleotides are those naturally occurring *Bordetella pertussis* sequences that encode allelic variants of the *Bordetella* pathogenicity proteins in *B. pertussis*.

20   The present invention further relates to polynucleotides that hybridize to the herein above-described sequences. In this regard, the present invention especially relates to polynucleotides which hybridize under stringent conditions to the herein above-described polynucleotides. As herein used, the term "stringent conditions" means hybridization will occur only if there is at least 80%, and preferably at least 90%, and more preferably at least  
25   95%, yet even more preferably 97-99% identity between the sequences.

Polynucleotides of the invention, which are identical or sufficiently identical to a nucleotide sequence of any gene defined in tables 2 and 3 or a fragment thereof, may be used as hybridization probes for cDNA and genomic DNA, to isolate full-length cDNAs  
30   and genomic clones encoding *Bordetella* pathogenicity polypeptides respectively and to

isolate cDNA and genomic clones of other genes (including genes encoding homologs and orthologs from species other than *Bordetella pertussis*) that have a high sequence similarity to the *Bordetella* pathogenicity genes. Such hybridization techniques are known to those of skill in the art. Typically these nucleotide sequences are 80% identical, preferably 90% identical, more preferably 95% identical to that of the referent. The probes generally will comprise at least 15 nucleotides. Preferably, such probes will have at least 30 nucleotides and may have at least 50 nucleotides. Particularly preferred probes will range between 30 and 50 nucleotides. In one embodiment, to obtain a polynucleotide encoding *Bordetella* pathogenicity polypeptide, including homologs and orthologs from species other than *Bordetella pertussis*, comprises the steps of screening an appropriate library under stringent hybridization conditions with a labeled probe having a nucleotide sequence contained in one of the gene sequences defined by tables 2 and 3, or a fragment thereof; and isolating full-length cDNA and genomic clones containing said polynucleotide sequence. Thus in another aspect, *Bordetella* pathogenicity polynucleotides of the present invention further include a nucleotide sequence comprising a nucleotide sequence that hybridize under stringent condition to a nucleotide sequence having a nucleotide sequence contained in one of the genes defined by table 2 and 3, or a fragment thereof. Also included with *Bordetella* pathogenicity polypeptides are polypeptides comprising amino acid sequences encoded by nucleotide sequences obtained by the above hybridization conditions. Such hybridization techniques are well known to those of skill in the art. Stringent hybridization conditions are as defined above or, alternatively, conditions under overnight incubation at 42°C in a solution comprising: 50% formamide, 5xSSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10 % dextran sulfate, and 20 microgram/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65°C.

The polynucleotides and polypeptides of the present invention may be employed as research reagents and materials for discovery of treatments and diagnostics to animal and human disease.

**Diagnostic Assays**

This invention also relates to the use of Bordetella pathogenicity polypeptides, or Bordetella pathogenicity polynucleotides, for use as diagnostic reagents. Detection of Bordetella pathogenicity polypeptides will provide a diagnostic tool that can add to or  
5 define a diagnosis of *B. pertussis* disease, among others.

Materials for diagnosis may be obtained from a subject's cells, such as from blood, urine, saliva, tissue biopsy.

- 10 Thus in another aspect, the present invention relates to a diagnostic kit for a disease or susceptibility to a disease, particularly *B. pertussis* disease, which comprises:
- (a) a Bordetella pathogenicity polynucleotide, preferably the nucleotide sequence of one of the gene sequences defined by tables 2 and 3, or a fragment thereof;
  - (b) a nucleotide sequence complementary to that of (a);
  - 15 (c) a Bordetella pathogenicity polypeptide, preferably the polypeptide encoded by one of the gene sequences defined in tables 2 and 3, or a fragment thereof;
  - (d) an antibody to a Bordetella pathogenicity polypeptide, preferably to the polypeptide encoded by one of the gene sequences defined in tables 2 and 3; or
  - (e) a phage displaying an antibody to a Bordetella pathogenicity polypeptide, preferably  
20 to the polypeptide encoded by one of the gene sequences defined in tables 2 and 3.

It will be appreciated that in any such kit, (a), (b), (c), (d) or (e) may comprise a substantial component.

25 **Vaccines**

- Another aspect of the invention relates to a method for inducing an immunological response in a mammal which comprises inoculating the mammal with Bordetella pathogenicity polypeptide or epitope-bearing fragments, analogs, outer-membrane vesicles or cells (attenuated or otherwise), adequate to produce antibody and/or  
30 T cell immune response to protect said animal from *B. pertussis* disease, among others. In

particular the invention relates to the use of Bordetella pathogenicity polypeptides encoded by the genes defined in table 3 – the effector proteins. Yet another aspect of the invention relates to a method of inducing immunological response in a mammal which comprises, delivering Bordetella pathogenicity polypeptide via a vector directing  
5 expression of Bordetella pathogenicity polynucleotide *in vivo* in order to induce such an immunological response to produce antibody to protect said animal from diseases.

A further aspect of the invention relates to an immunological composition or vaccine formulation which, when introduced into a mammalian host, induces an  
10 immunological response in that mammal to a Bordetella pathogenicity polypeptide (particularly one encoded by a gene defined in table 3) wherein the composition comprises a Bordetella pathogenicity gene, or Bordetella pathogenicity polypeptide or epitope-bearing fragments, analogs, outer-membrane vesicles or cells (attenuated or otherwise). The vaccine formulation may further comprise a suitable carrier. The  
15 Bordetella pathogenicity polypeptide vaccine composition is preferably administered orally or parenterally (including subcutaneous, intramuscular, intravenous, intradermal etc. injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the blood of the  
20 recipient; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing  
25 the immunogenicity of the formulation, such as oil-in water systems and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

Yet another aspect relates to an immunological/vaccine formulation which comprises the polynucleotide of the invention. Such techniques are known in the art, see for example Wolff *et al.*, *Science*, (1990) 247: 1465-8.

## 5    **EXAMPLES**

The examples below are carried out using standard techniques, which are well known and routine to those of skill in the art, except where otherwise described in detail. The examples illustrate, but do not limit the invention.

- 10    Example 1: A type III secretion system is present in a pathogenicity island in *Bordetella pertussis*.

The presence of a *lcrD* homologous gene in the *Bordetella pertussis* genome was investigated by polymerase chain reaction (PCR). The primers used (oligos 95080 and 95081 shown in Table 1) were degenerate oligonucleotides corresponding to highly  
15    conserved regions of the amino acids sequences of the LcrD/FlbF family of proteins. These primers were also designed to favour the amplification of virulence genes instead of their paralogue *flhA* or *flbF* flagellar genes, present in flagellated bacterial strains. The presence of the 3' triplet CAT in oligonucleotide 95081 is a determinant – indeed when multiple sequence analysis is done using known homologous sequences (database  
20    searching was done with either the FASTA and TFASTA programs of the GCG9 package, or with BLASTN, BLASTP and BLASTX programs, and alignments were carried out with the PILEUP program from the GCG9 package) it could be seen that the CAT triplet codes for a methionine which is exclusively present in virulence sequences while absent in the flagellar ones.

25

When analysed on agarose gel, the PCR product appeared as a heterogeneous mix of fragments, one of which was presenting the expected size (around 150 bp). A second round of amplification using the approximately 150 bp DNA as template yielded a single amplicon which was cloned in pCRII (obtained from Invitrogen) for further  
30    characterisation. It appeared as a 152 bp fragment whose nucleotide sequence (Fig. 1),

although similar to all *lcrD/flbF* homologous genes, shares a higher level of identity with the virulence (*lcrD*-like) genes.

5 **Table 1.**

oligonucleotides	sequence <sup>1</sup>	features	<i>lcrD</i> corresponding codons <sup>2</sup>
95080	GSH ATG CCW GGH AAR CAR ATG	direct, degenerate	150 to 156
95081	GC RTC DCC YTT DAC RAA YTT CAT	complement, degenerate	193 to 200
95363	CC ATC GAC GCG GAC TTG CGC G	direct, non-degenerate	157 to 164
95364	CGC GCC GTC CAT GGC GCC ATA	complement, non-degenerate	186 to 192
96110	C CGA CGC CGA CGC CGT ACG GTC	direct, non-degenerate	172 to 179

<sup>1</sup> The letter code for nucleotide ambiguity proposed by IUB (Nomenclature Committee, 1985, Eur. J. Biochem., **150**: 1-5) was used.

<sup>2</sup> The DNA sequence of the *lcrD* gene from *Yersinia enterocolitica* used for this work was published by Plano *et al.* (1991).

To ensure that the cloned fragment was actually a *B. pertussis* sequence PCR was performed under stringent conditions with serial 10-fold dilutions of DNA from *B. pertussis*. The optimisation of stringent PCR conditions require a perfect match between template and primers. It was likely, however, that due to the degeneration of the original primers, the 152 bp sequence initially obtained had, at its boundaries, a few base pair differences with the actual *B. pertussis lcrD*-like (hereafter called *bcrD*) sequence. A nested PCR approach using internal primers (oligos 95363 and 95364 Table 1) was therefore preferred, as primers known to be the correct *B. pertussis* sequence are used. A dose-response-relationship was observed between the 10-fold dilutions of *B. pertussis* template DNA and the product of the nested PCR, suggesting that the 152 bp amplicon actually originates from the *Bordetella* genome.

Comparison of the 152 bp sequence with *lcrD/flbF* genes allowed us to define a specific DNA stretch (oligo 96110 in Table 1) which was used as a probe for screening a genomic library of *B. pertussis* constructed in the plasmid vector pBR327 (Delisse-Gathoye *et al*, 1990, *Infect-Immun.* 58: 2895-905). Several positive clones were isolated and restriction analysis of their resident plasmids showed that they harboured overlapping inserts. The entire nucleotide sequence of one insert was determined, revealing a large open reading frame (ORF). This 2100 bp ORF encoded a 75 kDa polypeptide which is 59 % and 47 % identical to the yersinial proteins LcrD and FlhA respectively. Multiple amino acids comparisons of all known members of the LcrD/FlbF family of proteins, including the *B. pertussis* BcrD deduced amino acid sequence, showed that this sequence clearly ranked within the virulence associated determinants (Fig. 2). These data strongly suggest that *B. pertussis* possesses a type III export system, involved in the secretion of virulence effectors.

15

The *B. pertussis* *lcrD*-like nucleotide sequence (*bcrD*) has been submitted to EMBL and assigned the accession number Y13383.

This general technique has been useful for determining the presence/absence of a type III secretion system in other bacterial strains. The human pathogens *Borrelia burgdorferi* and *Helicobacter pylori* were intensively screened for such a system using this technique. No evidence for a type III secretion system could be found. The subsequent publication of the genome sequences of these microorganisms has confirmed the absence of similar systems in these species. In contrast, the method allowed the amplification of a DNA fragment from the phytopathogen *Pseudomonas corrugata*, which clearly ranks among the virulence sequences. This technique could be applied to any Gram negative pathogen of medical or agronomic importance such as *Neisseria* spp, *Moraxella catharalis*, *Vibrio cholerae*, any Enterobacteriaceae, *Pseudomonas* spp, *Haemophilus influenzae*, *Brucella* spp, *Francisella tularensis*, *Pasteurella* spp, *Legionella pneumophila*. Even in strains that have been fully sequenced, this technique

30

can be used as a simple method for checking alternate types or strains of the same species. For instance, some types of pathogenic *Escherichia coli* harbour a type III secretion system whereas others do not.

- 5 Example 2: Analysis of the *B. pertussis bcrD* flanking sequences to characterise the pathogenicity island and virulence-related proteins encoded therein.

The tendency for systematic clustering of type III encoding genes inside pathogenicity islands prompted the analysis of *B. pertussis bcrD* flanking sequences. The whole region containing the pathogenicity island was sequenced by chromosome  
 10 walking taking care to pay attention to the fact that each Pathogenicity island region must be represented in at least two independent clones, to avoid possible artefacts due to chimeric DNA inserts. This revealed clustered ORFs that could be classed in 3 categories: class I type ORFs (table 2); class II type ORFs (table 3) – the effector proteins which have the best vaccinal and diagnostic properties; & insertion sequences, and ORFs  
 15 homologous to house keeping genes of other species (table 4). Although there is no general rule for defining the boundaries of a Pathogenicity island, they can be demarcated with a direct or inverse repeat at one or other boundary, however the absolute demarcation of the boundaries can only really be done by the detection of house keeping genes at the extremes of the sequence. In the present case, an insertion sequence (IS in  
 20 Fig. 3) was present at the 5' end of the island (separating the virulence ORFs from the house keeping genes), but absent at the 3' end. In addition, the presence of house keeping genes (*greA* and ICFG-like) surrounding a locus which, according to sequence data, encompasses numerous virulence sequences is a good indication of the boundaries of the island. The complete gene organisation of the pathogenicity island is schematically  
 25 represented in Figure 3. The precise definition of the PAI boundaries requires further experimental data, such as the characterisation of the corresponding chromosomal region of a *Bordetella* strain which is devoid of a type III secretion system.



Table 2

names	Coding sequence from/to (with reference to Fig. 5)	Coding DNA strand	Homologous genes (from <i>Yersinia</i> , unless otherwise specified)
Class I genes, i.e. genes coding for determinants involved in the secretory apparatus and their regulation			
<i>bcrD</i>	8656/10755	complement	<i>lcrD</i>
<i>bcrH</i>	14097/14582	direct	<i>lcrH</i> (= <i>sycD</i> )
<i>bscC</i>	26955/28757	direct	<i>yscC</i>
<i>bscD</i>	7379/8659	complement	<i>yscD</i>
<i>bscE</i>	7039/7338	complement	none
<i>bscF</i>	6783/7049	complement	<i>yscF</i>
<i>bscI</i>	17892/18218	direct	<i>yscI</i>
<i>bscJ</i>	18215/19039	direct	<i>yscJ</i>
<i>bscK</i>	19032/19694	direct	none
<i>bscL</i>	19664/20302	direct	<i>yscL</i>
<i>bscN</i>	20307/21641	direct	<i>yscN</i>
<i>bscO</i>	21641/22150	direct	<i>yscO</i>
<i>bscP</i>	22147/22695	direct	none
<i>bscQ</i>	22692/23771	direct	<i>yscQ</i>
<i>bscR</i>	23768/24439	direct	<i>yscR</i>
<i>bscS</i>	24445/24711	direct	<i>yscS</i>
<i>bscT</i>	24723/25523	direct	<i>yscT</i>
<i>bscU</i>	25520/26569	direct	<i>yscU</i>
<i>bscV</i>	26566/26964	direct	none
<i>brpL</i>	28778/29380	complement	<i>hrpL</i> ( <i>Pseudomonas syringae</i> )

Table 3

Names	Coding sequence from/to (with reference to Fig. 5)	Coding DNA strand	Homologous genes (from <i>Yersinia</i> , unless otherwise specified)
Class II ORFs which putatively code for effector proteins			
<i>bopN</i>	11906/13003	complement	<i>YopN</i> (= <i>lcrE</i> )
<i>orf1</i>	6160/6747	direct	none
<i>orf2</i>	10752/11120	complement	none
<i>orf3</i>	11117/11527	complement	none
<i>orf4</i>	11532/11909	complement	none
<i>orf5</i>	13002/13784	direct	none
<i>orf6</i>	13806/14081	direct	none
<i>orf7</i>	14630/15571	direct	none
<i>orf8</i>	15601/16803	direct	none
<i>orf9</i>	16827/17288	direct	<i>bcrH</i>
<i>orf10</i>	17293/17814	direct	<i>pcr4</i> ( <i>Pseudomonas aeruginosa</i> )
<i>orf11</i>	29412/29591	complement	none
<i>orf12</i>	29555/30529	complement	none
<i>orf13</i>	30631/31776	direct	none
<i>orf14</i>	31773/33005	complement	none
<i>orf15</i>	32370/33014	direct	none

**Table 4**

No name specified	Coding sequence from/to (with reference to Fig. 5)	Coding DNA strand	Homologous sequences
Insertion Sequences and house keeping genes			
	711/2024	direct	uracil permease genes of numerous bacteria
	2055/3590	complement	chemoreceptor genes of numerous bacteria
	4220/4696	direct	<i>greA</i> ( <i>Escherichia coli</i> )
	4998/5948	complement	transposase genes of numerous bacteria
	33002/34852	complement	ICFG gene ( <i>Synechocystis</i> sp)

Next to the *bcrD* gene, there is an open reading frame (ORF) whose deduced amino acid sequence shares significant similarities with the YscU protein of *Yersinia* spp (39% identity and 51% similarity) and other known YscU homologs (Fig. 4). YscU, like LcrD, is a component of the *Yersinia* type III secretion machinery involved in the virulence mechanisms of the bacteria. *B. pertussis* therefore possesses a classical type III secretion system which is most probably involved in pathogenicity. This latter point can be investigated through phenotypic analyses of mutants (see below).

The total length of the Pai is approximately 30 to 40 kb. The DNA sequence of the whole region is presented in Figure 5, and is referred to in tables 2, 3, and 4.

No homologies could be found between the *B. pertussis* Class II Pai DNA sequences and the sequences reported in the GenEMBL databases (except for those stated in table 3). The expressed products of these unknown genes within the Pai responsible for virulence, will be useful in the development of a vaccine formulation against pathogenic *Bordetella pertussis*.

To address the precise function of the *Pai*, a *bcrD* mutant was engineered by allelic exchange. In the resulting mutant, the *bcrD* gene was disrupted by an *aphA-3* cassette conferring kanamycin resistance. This cassette was inserted in such a way that translation was not interrupted, avoiding any polar effect on expression of putative downstream cistrons. A mutant has been isolated and its associated phenotype is being  
5 currently analysed.

## Bibliography

- Allaoui, A., Woestyn, S., Sluiter, C., and Cornelis, G. R. (1994) YscU, A *Yersinia enterocolitica* inner membrane protein involved in Yop secretion. *J Bacteriol* 176:4534-4542.
- Bergman, T., Erickson, K., Galyov, E., Persson, C., and Wolf-Watz, H. (1994) The lcrB (yscN/U) gene cluster of *Yersinia pseudotuberculosis* is involved in Yop secretion and shows high homology to the spa gene clusters of *Shigella flexneri* and *Salmonella typhimurium*. *J Bacteriol* 176:2619-2626.
- Bogdanove, A. J., Wei, Z.-M., Zhao, L., and Beer, S. V. *Erwinia amylovora* secretes a Harpin via a type III pathway and contains a homolog of yopN of *Yersinia* spp. *J Bacteriol* 178:1720-1730.
- Fenselau, S., Balbo, I., and Bonas, U. (1992) Determinants of pathogenicity in *Xanthomonas campestris* pv. *vesicatoria* are related to proteins involved in secretion in bacterial pathogens of animals. *Mol Plant Interact* 5:390-396.
- Gyri (1995). *Mol. Microbiol.* 15: 761-769.
- Hueck, C. J. (1998) Type III protein secretion systems in Bacterial Pathogens of Animals and Plants. *Microb. Mol. Biol. Rev.* 62: 379-433
- Plano (1991). *J. Bact.* 173: 7293-7303.
- Ramakrishnan, G., Zhao, J. L., and Newton, A. (1991) The cell cycle-regulated flagellar gene flbF of *Caulobacter crescentus* is homologous to a virulence locus (lcrD) of *Yersinia pestis*. *J Bacteriol* 173:7283-7292.
- Yuk, M. H., Harvill, E. T., and Miller, J. F. (1998) The BvgAS virulence control system regulates type III secretion in *Bordetella bronchiseptica*. *Mol. Microbiol.* 28:945-959.

**We claim:**

1. A method of identifying virulence genes from a pathogenicity island containing a type III secretion system from pathogenic strains of bacteria, comprising:
  - 5       designing degenerate PCR primers complementary to well-conserved regions specific to the LcrD polypeptide of *Yersinia*;
  - amplifying the polynucleotide containing the DNA sequence between (and including the DNA sequence of) the primers of *lcrD*-like genes present in said pathogenic strain of bacteria;
  - 10       sequencing the *lcrD*-like gene;
  - determining whether the DNA sequence is more homologous: to the virulence-associated family of *lcrD*-like genes, or to the flagellar-associated family of *lcrD*-like genes; and
  - if a virulence-associated member, sequencing the entire pathogenicity island, and
  - 15       identifying genes within this sequence.
  
2. A method of determining whether a particular bacterial strain harbours a type III secretion system involved in pathogenicity, comprising:
  - designing degenerate PCR primers complementary to well-conserved regions
  - 20       specific to the LcrD polypeptide of *Yersinia*;
  - amplifying the polynucleotide containing the DNA sequence between (and including the DNA sequence of) the primers to determine the presence of any *lcrD*-like genes in said bacterial strain;
  - if amplified successfully, sequencing the *lcrD*-like gene; and
  - 25       determining whether the DNA sequence is more homologous: to the virulence-associated family of *lcrD*-like genes, or to the flagellar-associated family of *lcrD*-like genes.
  
3. An isolated polynucleotide comprising a nucleotide sequence encoding the BopN
- 30       polypeptide of *Bordetella pertussis*.

4. An isolated polynucleotide comprising a nucleotide sequence encoding the Omp1 polypeptide of *Bordetella pertussis*.
- 5 5. An isolated polynucleotide comprising a nucleotide sequence encoding the Omp2 polypeptide of *Bordetella pertussis*.
6. An isolated polynucleotide comprising a nucleotide sequence encoding the Omp3 polypeptide of *Bordetella pertussis*.
- 10 7. An isolated polynucleotide comprising a nucleotide sequence encoding the Omp4 polypeptide of *Bordetella pertussis*.
8. An isolated polynucleotide comprising a nucleotide sequence encoding the Omp5 polypeptide of *Bordetella pertussis*.
- 15 9. An isolated polynucleotide comprising a nucleotide sequence encoding the Omp6 polypeptide of *Bordetella pertussis*.
- 20 10. An isolated polynucleotide comprising a nucleotide sequence encoding the Omp7 polypeptide of *Bordetella pertussis*.
11. An isolated polynucleotide comprising a nucleotide sequence encoding the Omp8 polypeptide of *Bordetella pertussis*.
- 25 12. An isolated polynucleotide comprising a nucleotide sequence encoding the Omp9 polypeptide of *Bordetella pertussis*.
13. An isolated polynucleotide comprising a nucleotide sequence encoding the Omp10 polypeptide of *Bordetella pertussis*.
- 30

14. An isolated polynucleotide comprising a nucleotide sequence encoding the Omp11 polypeptide of *Bordetella pertussis*.
- 5 15. An isolated polynucleotide comprising a nucleotide sequence encoding the Omp12 polypeptide of *Bordetella pertussis*.
16. An isolated polynucleotide comprising a nucleotide sequence encoding the Omp13 polypeptide of *Bordetella pertussis*.
- 10 17. An isolated polynucleotide comprising a nucleotide sequence encoding the Omp14 polypeptide of *Bordetella pertussis*.
- 15 18. An isolated polynucleotide comprising a nucleotide sequence encoding the Omp15 polypeptide of *Bordetella pertussis*.
19. An isolated polypeptide encoded by the polynucleotide of claims 3-18.
20. A vaccine comprising the polypeptide of claim 19.
- 20 21. A kit for diagnosing infection with *B. pertussis* bacteria in a human comprising a polynucleotide of claims 3-18 or a polypeptide of claim 19.



```

1  GGAATGCCAGGAAGCAGATGTCCATCGACGGCGGACTTGCGCGGGGCACCATAGACATG
   +-----+-----+-----+-----+-----+-----+
60 CCTTACGGTCCTTTTCGTCTACAGTAGCTGGCCTGAACGGCGGCCCGTGGTATCTGTAC
   +-----+-----+-----+-----+-----+-----+
   G M P G K Q M S I D A D L R A G T I D M -
   +-----+-----+-----+-----+-----+-----+
61 GACGAAGCCCGACGGCGACGGCGGTACGGTCGAGAAGGAAGCCAACTGTATGGCGCCATG
   +-----+-----+-----+-----+-----+-----+
120 CTGCTTCGGGCTGCGGCTGCGGCATGCCAGCTCTTCCTTTCGGTTGACATACCGCGGTAC
   +-----+-----+-----+-----+-----+-----+
   D E A R R R R R T V E K E S Q L Y G A M -
   +-----+-----+-----+-----+-----+-----+
121 GACGGCGCGATGAAATTGTCAAGGGCGACGC
   +-----+-----+-----+-----+-----+-----+
152 CTGCCGCGCTACTTTAAACAGTTCCTCCGCTGCG
   +-----+-----+-----+-----+-----+-----+
   D G A M K F V K G D A -

```

Fig. 1

**THIS PAGE BLANK (USPTO)**

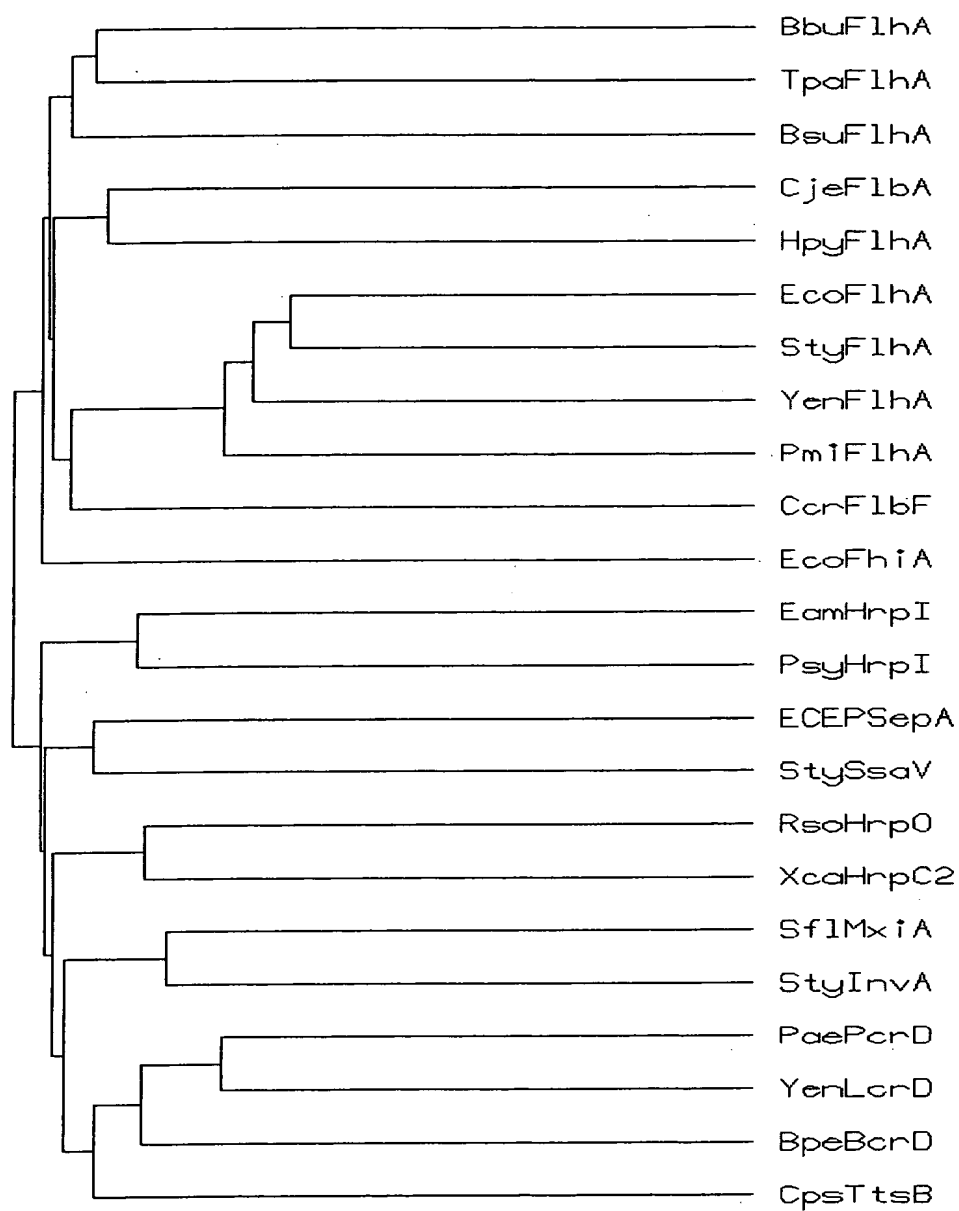


Fig 2

**THIS PAGE BLANK (USPTO)**



— 1 kb

Fig 3

**THIS PAGE BLANK (USPTO)**

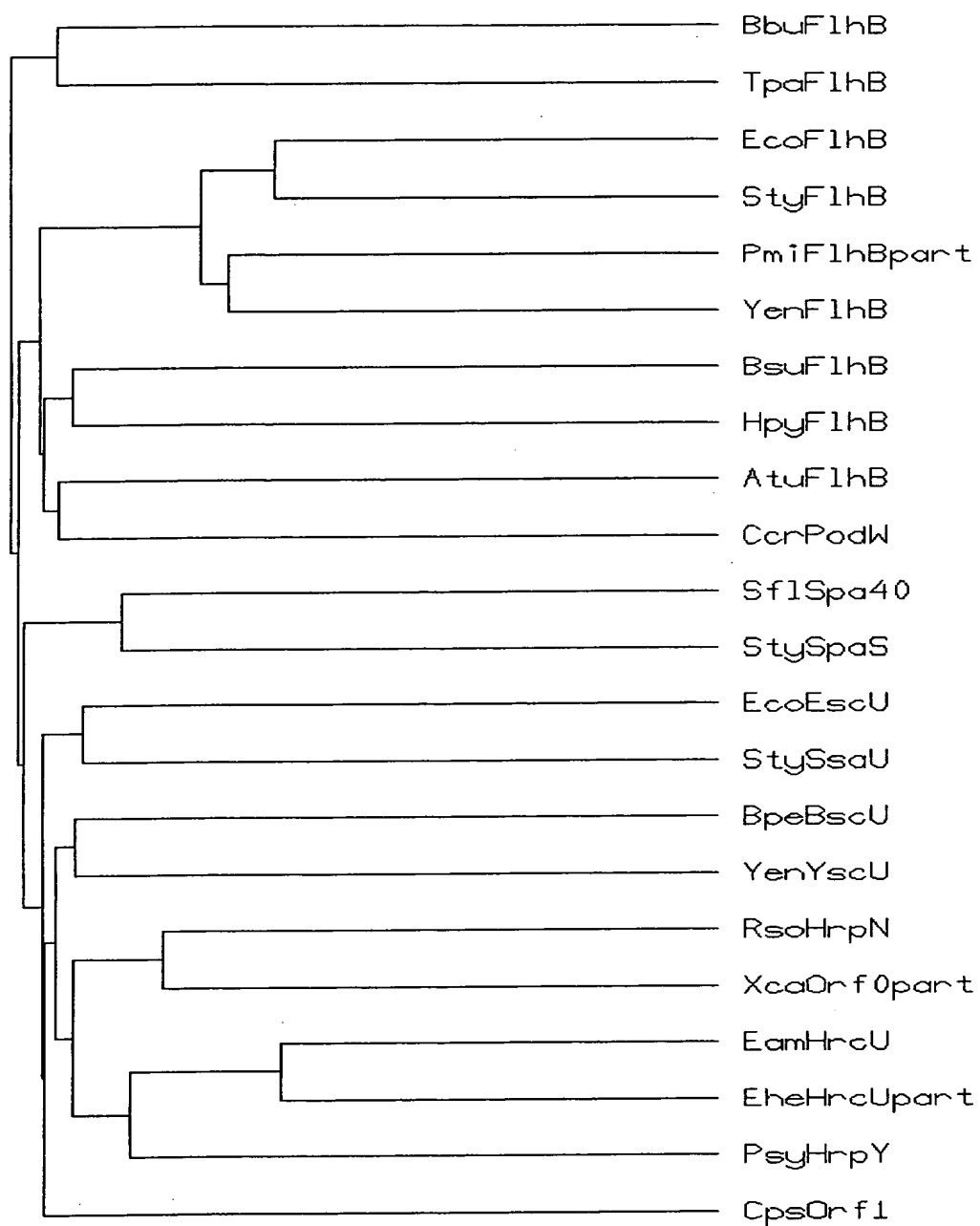


Fig. 4

**THIS PAGE BLANK (USPTO)**



1 GATCTCCAGC TTGATGTCCG GATGGGCCTT CTCGAAGGCC GCCTTGTAGG  
 51 CCTGGGTCAG GTCCTTGGGG AAAGACGTGA TGACCGTTAC CGTGCCGGCC  
 101 AGGGCCGGCG CGCAGGCAGC CATGATGGCC GCTGCGAGGA CTGGGCGCAA  
 151 CGCTTGCATG GGTCTCCTCC TTTCTTGAGT TGTGGCAAgA CCTTAATGCC  
 201 CGTGTTCCCC CGAGTCCAAT CAGAAATTTT GATGCGCGCC ATCaACgGCG  
 251 CGGCcGGCTT CAGCGCGCCG CCAGCCGCTG CGCCAGGCGC AATTGCTCTT  
 301 CCAGCAGCTT CTGGATCTGC CCCTGGATGG CGAAGGCCTG GCGGTTCAGG  
 351 TCCTGCAGCT GCCGGGCTG CTGCTCGGCG GTGGCGTTGC TTTCGCGCAC  
 401 CTGGCGCATC TGGCGCATGA TGGCCGCCAG CTGGCGCTGC AGGGCCTGCA  
 451 AGGCCTCGGC CGCGGGGCTG CCGGCCTCGT CCTTTGCGCC GCTGGCGGCG  
 501 GCCCGGTGCG GCAACTGCCT GGCGTACTGG TCGAACGCGC CGGCGGAGAT  
 551 GCGCGATGAA ACGGAAGTGA CGGTCATGGT GGCGGCGTCC GCGTGGATTA  
 601 GGGTTAGTCC TTAATACGGC AGGCGCGCGG ACGAACTTAA ACATATTGCT  
 651 GCGCGCTTCC GACCCCCCTT AACTTGC GC CGCCATCCC TTCAACTTCC  
 701 ACGCATACCT ATGTCCAATA CCTATTCCC GCGCTGGCGG CTGGCCGACG  
 751 ACACCGTGCC GGGCGCGGTC ATCGCGCCCG ACGAACGCCT GTCCTGGCCC  
 801 AAGAACATCG CCATGGGGGC CCAGCACGTG GTCGCCATGT TCGGTTCCAC  
 851 CGTGCTGGCG CCGTGCTGA TGGGTTTCGA CCCCAATGTG GCGATCCTCA  
 901 TGTCCGGCAT CGGCACGTG ATCTTCTTCC TGTTGTCGCG CGGCCGGGTG  
 951 CCCAGCTACC TGGGCTCCAG CTTGCGCTTC ATCGGCGGGG TGGTGGCGGT  
 1001 CACCGGCTAT GTGGCGCCCC GCGCCAACGC CAATATCGGC GTGGCGCTCG  
 1051 GCGCGATCAT CGCCTGTGGC CTGGTGTACG CGCTGATCGG CCTGGTCGTA  
 1101 TGGGCGGCCA GCGCGCGCGG CAACGGGGCG CGCTGGATCG AGGCCATGAT  
 1151 GCCCGCGGTC GTCACGGGCG CGGTGGTGGC GGTGATCGGC CTGAACCTGG  
 1201 CCCCgATCGC CGCCAAGGGC GCCATGGGTT CGTCCGGCTT CGAGGCCAGC  
 1251 ATGGCGTTGA TGACCATCCT GTGCGTGGGC GGCATCGCCG TCTACACGCG  
 1301 CGGCATGGTG CAGCGGCTGC TGATCCTGGT CGGCCTGGTG CTGGCCTGCG  
 1351 TCATCTACGC GGTCTGCGCC AACGGCCTGG GGCTGGGCGC GCCCATGGAC  
 1401 TTCGCCAAgG TGGCCGCCGC GCCGTGGTTC GGCTGCCCCA GCTTCGCCGC  
 1451 GCCGGTGTTC GAgCCGCAGG CCATGGGCCT GATCGTGCCG GTGGCCATCA

**Fig. 5** 1/23

**THIS PAGE BLANK (USPTO)**

1501 TCCTGGTGGC CGAGAACCTG GGCCACGTGA AGGCGGTTCG CGCCATGACC  
1551 GGACAGGACC TGGACCGCTA CGTGGGCCGC GCCTTCGTGG GCGACGGCGT  
1601 GCGGACCATG GTTCCGGCG CCGTCGGCGG CACCGGGGTG ACCACCTACG  
1651 CCGAGAATAT CGGCGTGATG GCCGTGACGC GCATCTATTC CACGCTGGTG  
1701 TTCGTGGTGG CGGCCGTGAT CGCGCTGGTG CTGGGGTTCT CGCCCAAGTT  
1751 CGGCGCGCTG ATCCAGACCA TCCCCGGCCC CGTGCTGGGG GGCATGTCGG  
1801 TCGTGGTGTG CGGCCTGATC GCCATCGCCG GCGCGCGCAT CTGGGTGGTC  
1851 AACCAGGTCG ATTTTCAGCGA CAACCGCAAT CTGATCGTGG CCGCCGTGAC  
1901 CCTGGTGCTG GGGGCGGGCG ACTTCAGCGT CAAGCTGGGC GATTTCTCGA  
1951 TGAACGGCAT CGGCACCGCC ACGTTCGGCG CCATCATCCT GTACGCCCTG  
2001 CTGGGCCTGG CGCGTCGCCG CTGACGGCGC GCCAACCCGG CACcGGGGCC  
2051 GCGCTCAGTG GGCCGCGGCC GGCcGcAGGG CGGCATGGCC GCCCGCCGGC  
2101 CCGGCCAGCG GCGCGGCGGC GACGTCGATG ATGGCGCCGC CCTGCGTCCT  
2151 GAAGACCGCC GCCGCGGCCG CCAGGTGGCG GGCTGTTCT TCCATGGGCC  
2201 TGGCCCGCGC CGCGGCCTGT TCGACCAGCG CGGCATTCTG cTGGGTGGTT  
2251 TCGTCCATTT GCGAAATCGC CAGGcTGACC TGGTCGATGC CGCTGGCCTG  
2301 CTGGGCCGAG GCGGCCGAGA TCTCCCCCAT GATGTCGGCC ACGCGCTGTA  
2351 CCGAGGCCAC CACCTCGTCC ATGGTGCCGC CGGCGCTGGC GACCTGTTGC  
2401 GAGCCGGCGC GCACCGTCGC CACCGAGCTC TCGATCAGGG CTTGATCTC  
2451 CTTGGCCGCC TGCGCGGCGC GCTGGGCCAG GCTGCGCACC TCGCCCGCCA  
2501 CCACCGCGAA GCCCTTGCCC TGTTGCGCCG CGCGCGCCGC cTCGACCGCG  
2551 GCGTTCAGCG CCAGGATATT GGTCTGGAAG GCGATGCCGT CGATCACGGT  
2601 GACGATGTCG GCGATcTGGC GCGAGCTCGC GGAGATGCCG TGCATGGTCT  
2651 GCACCACcTG CGCGACCGAC TCGCCGCCGC GCTGCGCCAC cTGCATGCTG  
2701 ACGGCGGCCA GCTGATTGGC CTGCGCGGCA TTGTCGGCGT TCTGCTTGAC  
2751 cGTGGTGGCC AGTTCcTCCA TGGTGGCCGC GGTcTCTTCC AGGGcGGCGG  
2801 CCTGCTCCTC CGTGCGGCTG GACAGGTTGG CGTTGCCGgC CGAGATCTCG  
2851 GCCGCGCCGA CGTTGATTTC GTCGACGCCG CGCCGCATGA CGGCGATGGT  
2901 GCGCGTCAGG CCTTCCTGCA TGCGCTTGAG CGCCGCGAAC AGCGCGCCGA  
2951 TTTCATTGGC CGAGCGCACC TCGATGCGCG CGGTGAGGTC GCCGTGGCGG  
3001 ATGCGGTCTGA AATGATGGCC GGCCTCCAGC AAGGGGCGCA GCACCGCGCG

Fig. 5 2/23

**THIS PAGE BLANK (USPTO)**

3051 GCGCACGAAC AGCCAGCCGG CCAGGGTCAG CAGCACGCCC AGCGTGGTCA  
 3101 GCGCGATGGC GCTCCAGCGG GCCACGACAT GGGTGTCTC gGnGCCGCTG  
 3151 CGCACTTCGT CGCTGTGCGC CTGTACCCTG GCCAGAAAGG CTTCCATGTC  
 3201 GCGCTGGAAT GCATGTTTCGG CCTGCTGCGC GCGATGCATG GCGGCCAGGG  
 3251 CGGGCTCgGC CTGgCCGGCG TCGACGGCGG CGGCCAGCTC GTCCAGCACC  
 3301 GACTGGTAGG CGCGCCAGCG CGTCTGCAGG GTCGCGGCCA GCTCGGCCG  
 3351 GGCCGGCTCC TTGGGCACGT CCACGTAgCG CTGGAATTGC gTGGCCGCT  
 3401 GCGCTGCGC CGCGCCGCTG GCCTCGAACA GCGGGTCGAC CTGgTTGACG  
 3451 GCgACCTGGT TCAGCCCCCTC GATCTCGGCG GCGCTCCTGC CCGCGTTGCG  
 3501 CCACGTCAGG GCGcCCGACA GCATCAGCAC CGCCAGGAAg CACGCGAATA  
 3551 CCGCCACCAT CGCGGTGCGG ACCTTGATCC TTTCAAgCAT GGGCATCTCC  
 3601 TGgGTAGGGT CTTGCGCAAT AAACGTCGAC GACGTGCGCG GCGTACCCG  
 3651 GCGGGAAGTG TCAGGCCGTG ACGTGAATC CGTTGTTGAT GTGGAATTG  
 3701 ACTTAAATGA TCGGTATTCG GTATTAATTT GTTATTTGTA GTTATATATA  
 3751 CGAATATTCA TACCCGGATT TGCCCTAAGT TGGTGCGTTC TCGACGGGTG  
 3801 cTTTcGATTG CCCGGgGCcT GCGGCGCATG AAGGAATGcG CGGCAACgCc  
 3851 GGGCCCGCGC TGCGATCGGC GGCCAACACG CAGGTTTTGT GGCTTTTCCG  
 3901 CAGCCAACAT GCGCCGAAAC CTACGCCGGG CTTACAGGCT TGCAATTCCG  
 3951 GTGGACTTTG CCGACAATGT CATCTGATTG CCCCGGTTCC GACCCGAGCC  
 4001 GGGGTTTTGT TTTGGTCGAC GCTTGGCCGC CGGATGCGGC AGGCCGATCA  
 4051 AAGAGGAGAC AGCAAAGGGG AGCCTCGGTC GGGTTCGACC TTGTCTCGTC  
 4101 TTTTGTTCGC CTGTCTCCG CAGCGGCCCG CCTGTTTCAT GGCGAGGCCA  
 4151 TACACCAAGC CGAGACCTTC ACCGAAACGC TCCGTCGGGG GCGTTTTCTA  
 4201 CTTTTGTTG GGAAACGACA TGTCTGCCAT TCCTTTGACC GTGCGCGGGG  
 4251 CCGAGCGCTT GCAGCAAGAA CTGCATCGGC TTAAGACCGT TGAGCGTCCT  
 4301 GCGGTGATCA GCGCCATTGC GGAGGCGCGT GCGCAGGGTG ATTTGTCGGA  
 4351 AAATGCCGAG TACGACGCCG CCCGCGAACG CCAGGGCTTC ATCGAAGGCC  
 4401 GGATCTCCGA ACTCGAGGGC ACGCTTTCGA ACGCGCACCT CATCGATCCA  
 4451 ACGGCGCTCG ACGCCGAAGG CCGTGCCGTG TTCGGCGCGA CCGTGGAAT  
 4501 CGAAGACCTC GACTCGGGCG ACCGCCTGAC CTACCAGATC GTGGGCGACG

**THIS PAGE BLANK (USPTO)**

4551 TCGAAGCCGA CATCAAGTCC AACCTGATTT CGGTCTCCAG CCCGGTGGCC  
 4601 CGCGCCCTGA TCGGCAAATC CGAGGGCGAT GTGGTCGAAG TGAAGGTGCC  
 4651 GGCTGGCGTG CGCGAGTACG AAGTCATCGG TGTGCGTTAT CTCTGACGCC  
 4701 GATTCCGCCC CCCTGCATAC CCATGGCCAA TGACCGACGC CGTTTCCACC  
 4751 ATCAGCCAGC CGCTTGCGGC CAGGCGGGGT GGCACAGCCG CAGCCGTCGC  
 4801 GCTGCAGGCG TCCCGGATGC GATATCCCGG GCTCCGGTCG TGGCCAGCGT  
 4851 ATTCCCCTTC TCGGCCGTCA ACGGCTTGGG GGGCGCGGTA TTCATCCGAA  
 4901 TTCCTGAACG CCGGGCCGTG CGCCGGGCGT CTGTATCGTT GTCGCGCCGC  
 4951 GATGCACGGA ACGTGCCGTC GGGCGCCGCA ACGCCGGGCC GCGCCGCCTA  
 5001 GGTGTGAAC TCAATAGGT TGTATTGTC CAGGTTGAGT CTGGAGATGG  
 5051 GTACAGCGCG CCCGATGCCT TGGTGGGGTC GATGCCAGTT GTAGTGGTGT  
 5101 AGCCAGGATT TCATGGCATC GGCTCGGTGT TGGGAGTTCT GGTA<sub>g</sub>GTGTG  
 5151 AGCGTAAGCC CACTCACGCA AGGCCGACTG GATGAAGCGT TCGGCCTTGC  
 5201 CATTGGTCTG TGGGCGGTAA GGTCGGGTAA AGCGGTGCTT GATGCCCAGC  
 5251 TCATGGCACA gCGCGGCGAA nGCGCgGCTG CgAAAgGCCG AGCCATTGTC  
 5301 GGTGAGCAAG CGCTGGATGG TCACGCCCAG GCGCTGGTAG TAGGCCACTG  
 5351 CGTCCTTGAG GAACTGGACG GCGCTGGGGA AGCGCTCGTC GGGGTGGATG  
 5401 TCGGTGAAGG CCACGCGGGC GTGGTCATCG ATGGCCACGA AGACGAAGTC  
 5451 CCAGCCGGCC CCCTCAACGG TATCGCGTCG GTTGCCCGTG ACCCGGTGGC  
 5501 CAGGGCGCTG GATACGTCCC AGCTTCTTGA TGTGATGTG CAGCAGATCG  
 5551 CCGGGGGCCT GATGCTCGTA GCGCACCACC GGCTCGGCCG GCTCCAGGTC  
 5601 GGCCAGGTGC GACAGACCGG CGCGGGCCAG GACGCGGCTG ACGGTGCTGG  
 5651 CTGACACGCC CAGCGCCTGG GCGATGCGCG CTTGGGTCAG CCGCTTGCGG  
 5701 CGCAGCTCCA CGATAGCCAG CGCCTTGGCC GCGCAATCG CTCGGGGCGA  
 5751 GACCGTCGGG CGCGAGGACG CATCGGCCAA GCGCGCTGG CCCTGAGCCA  
 5801 GGAAGCGGCC CAGCCATTTG CGCACAGTCG GCGCGGTGAC CCCATAGGCG  
 5851 CGGGCCGCTT CAGGCACACA AACTTGATGG GCGATCAATT GCTGGACCAT  
 5901 TTCGAGTCGA CGTAGGAAGG TCAATCGGGC ATGCTTATGG GTGTTTCATCC  
 5951 GGCCGGGCTC CTTGAGTGAA CTGGGGGGgT GGCGATTTCC AGTTTCTCAA  
 6001 ATCCGGTTCG GATGAACCAT GCATACAACC TATTGAATCT TCACAAC TAG  
 6051 CGCGCGTGGC GCGGAAAGAC CAGCAGGTCG GCCGTCACCG GTTCCCTGTT

Fig. 5 4/23

**THIS PAGE BLANK (USPTO)**



6101 GTCGAATAAC AGGTAATGCG TGGCGTCCGG ATATTGCCA TGGCGCGCGC  
6151 CGGATCCGCA TGCCCTTGGC GGGGTCTGC GGCCGGACGT CCGAAAACCG  
6201 GGATATTCCG AGAAGCAATC GGCTGGCGCC TTCCAGACTG TGCGCATACC  
6251 ATTGCCCTCT TTTGCCACGC ATTTGAGCG TATGGTTTCC CTTGCGCCCG  
6301 ACGCAGGCTC GCCTCGCCAT GACCGACACG GCATACCACC AACTCATCGC  
6351 CGATTTTCGGC CGCCTCATCG GCATCGACTC GCTCAACCCC GGTGCCGGCG  
6401 GCCTGTGTCA GTTGATTTTC GAACCGTGCG CACCGGTCTT CATCGCACCG  
6451 GTGCACGCCC GGACGGAAAT CATGATTTCC TGCGTGCTGG GCACGGCGGA  
6501 CGCGGCCAAC CCGGCAAGCA TGGCCCGAGC CAACTTCATG CAGGCCGGCA  
6551 GCGGCGTCGT GGCCTGCATC GGCGGCGATG GGTGTTCTA TCTGCAGCAG  
6601 GCCATACCCC TGTCGCGCGC CACGCCCGCA ATCCTGCTCG ATCACTGTGA  
6651 GCGTCTGCTG CAGGAAGCCT CGCGCTGGCG CGTCGGCGAC CACGACGGCT  
6701 GCGCCACCTC GGCCCCGAAT ATCGCCGCGC TGACGCGCGG CGTCTAGGTC  
6751 GGCGGCGACT GCCGGGCGCC GCCGCGGAG GCTCAACTCG CcTTCTGTAT  
6801 GACGCCCTTG AGCGAATCCG CGACCTGCTT GACCACCGTG CTCTGTAGAT  
6851 CGATCATCAC GACCCACGAT TGCATTTCTT GTTGACGAT CAGCAGGTGC  
6901 GACGTGCTGA CCGCGCCGTC TCCGCGCGCG CTGAGCGCCT CCAGGCGGCT  
6951 GCGCAGGTCT CGTTCGTGAG CGTTCAGCCG CGTATTGACC GCCTGGTTGA  
7001 CGCTCTGCAT GGTCACTCGG CCTGCGTCGC CTCCCAGGTT AATGGCCATG  
7051 CTTGTCTCCT TCGGCGCATT GTTCATTGCT CAGGCGCGTC AAGACTGACG  
7101 CCGGAGGGTT GTCCGGCCCG GTCGGGCGCT GCAGCAATAC CTGCCGGGCC  
7151 GCGGTGGCGG CCGCCAGCGC GTCGCGCCAT ACGGGATAGG TGTCGCGCCC  
7201 CACGCCATCG TGCAGGCGGA CGCGCAAACG TGTCTCCAGT TCGCCAAGCT  
7251 GCGACAGCAA GGTGTCGCGC AAGGCGGAAC CGCCCGGCGA TGCCAGGCGC  
7301 ACTTCCAGTT CGGTCAGGGC TAGAACAGAT GTACTCATAT GATGTTGCAG  
7351 ACGAGGGTTG ACGGCTACCG AGGCGATTTC ATCGCGTCAC GATGACCGGT  
7401 TCGGGACCAT CGAAGACCAG GCGGCCGGAT TCGATGGCGG TAAGGCGATA  
7451 CTGGTCCCGC AGTCCGCCCA CCAGGAGGCG GCTGCCATCG GCCAGCATCA  
7501 GGTACGGTTG CGGGCCGCTC ACGACACTGC GTATCTCGAA CGGCACGTGA  
7551 TCCCGCGTCG CGCGCGCGGC GGTGGCCGGC AGCCGGACGA CGTCGTAGTT

**Fig. 5** 5/23

**THIS PAGE BLANK (USPTO)**

7601 GCGCTGGTTG AACGCGGCCA CCAGCTCGCG CAGGCGCGCC ATGCGGCCTG  
 7651 CCGCCAATCC GCCGGGATCT GCGTCCAGGC GGTCGGCGTG CCAGCTGAGC  
 7701 TTGACCCGT CGAGGCGTTC GTCGGCCAGC TGGGCCGCGA ACTGGGCCGA  
 7751 GACCTCGTCG GCCAGGCGTA CATCGCGACC GAGGATCGTC ATGCCCCGCA  
 7801 GCGCATGCG CACCGCATGC AGCGCCGCGG CGCGTTCGTG CGCATCGCTG  
 7851 GCGATGCCCC AGATCGCCAG GCGGCCATTG CCGTACGGGC GCGCCATGTA  
 7901 GCGCACCCCG AATGTCGCCA GGACATCGCA GGCCAGGGCC CTGGCCTCGT  
 7951 CCTGCCTGCT TACCTGCATG GCAGGCCGTG GCGCAAGTTG CGCCAACGCC  
 8001 CTGGCGACCC GAGCGAATTC CGTCTCGTCG TGCACCCATC CGGTCACGGT  
 8051 GAGCACGCCG CCACGGCCGT AGGCCGCTTG TAATTGCTCG GTAAGGCCCA  
 8101 GGCTGTCGAT GAGCGCCGCG GCGCGGACCA GCGGCGCGGT GGGCGTTGGG  
 8151 GCGGCGCGCG CCGGCGGCGT GCGGGGTGTG GTCACGGAAA CCAGCGCCGT  
 8201 GGCCAGGCCG ACCAGCAGGA CGGCCGCGGC CGCGCCCAGC GCCAGCCAGG  
 8251 GCCGTCTGTC ACGTCGGCGC GGCATGAGGG CAGCGACGGA CGGCGGGCTT  
 8301 GTCGAGCCAG GGACGTCGTG CAAGGCTGTG TCGCTGCCGT CCGGGCCGCA  
 8351 CGGCTCCGGC GCGCGGGGCC ACGGCGCGGA AGGGGCGGCC ACGGTGATCC  
 8401 AGGCGGCTCC CAGCTCTACG GGTTCGTTGA AGGCCGCGGG CGGACACGGC  
 8451 GCCTGGGCGT CCAGGCCGGG CGTCACGGCG CCGGCCAACC GCCAGCCGGA  
 8501 CTGGTCGATC TCCAGCCATC CCGCCACTTC GGGCATGTCC TCGCCGGTCA  
 8551 GGACGATATC GCAATGCGGA TTGGCGCCCA CGCGCGCGCC ATGCACGGCC  
 8601 GGGCAGCGCG CCATGCACTG TCGCCTGAA AGCACGCGGA ATTCCAGCGC  
 8651 CGTCGTCATA GATCCACCCT GCCCAGGGGC TGTACATTGA TCTCCGGCGT  
 8701 CAGTTCCTGG TAGGACAGCA CCGGCAGGGC GTAGAGATCG GCTTCTATCA  
 8751 TCTTGCGCGT GTAGCGCCGG ATGTCCATCG ACGTCAGCAA GACGGGACGG  
 8801 CTCGCGCCGG CGGCCAGATC GCCGACACAT TGACGGATGT GCTCGACCAG  
 8851 TCGGCGTGTC GTGTCCGAT CGAGGGCGAG ATAAGTCCG GCGGCGGTCT  
 8901 GCCGGATGGC GCGCGCACG GTTTCCTCGA CCTTGGGGGC CAGCAGGTAG  
 8951 GCGGGCAGGA TATTGTGGCC GCTGGTGTAC TTGTGGCTGA TATAGCGCTT  
 9001 GAGTGCATT CGGACATACT CCGTAAGCAG GACGGTATCC TTTTCCTTCT  
 9051 GGCCCCATTC GACCAGCGCT TCCAGGACGG CGCGCAGGTT GCGTATCGAC  
 9101 ACTTCTTCGG AAACAAGGCG CTGCAGGATT TCGGCAATCT TCTGCACCGG

Fig. 5 6/23

**THIS PAGE BLANK (USPTO)**

9151 CATGACGCGC AGGCACTCCT TGACCAGATC GGGAAATCGT TCTTCCATGG  
 9201 CCGAAAGCAG AAACCGGGTT TCCTGGATGC CGATGAAATC GGCTGAATAT  
 9251 TTTTTC AATA CATATGCCAA GTGCCAAGTC AGGATCTGGC TGATACCCAG  
 9301 GTAAGGAATA CCTGCATCGC GCAAGGCGCC GGTCAGACTG GCCGCAACCC  
 9351 AGATCGTGGG CGTATCGGGC AGAAAGGCCG CGCCCGTTTC GTATGCGATC  
 9401 CGCAGGGCCT GCAGGTTCTG CTCGGTGTCC CGCACCAGCA CGGCATCGTC  
 9451 GCGCAACATT CCTTGCGCCA CCGGGATCTC CGACAGCACG ATGGTGTAGG  
 9501 TATTGGCGGC CAGCGCTTCG GTGAAGCGCA ACTGGATGCC GGGAAACGGC  
 9551 ACGCCAGGT CGAAATAGAG CGCCCGCCGG ATCTGCAGCA GATCGTCGGT  
 9601 GAGGGTGGCC GGCTCGAACC GGGGCTGCAG CCGCGCGGCT ACGTCGATGA  
 9651 TCAGCGGGAC GGTGGGGGCG AATTCCGCCT GCCCATCCGC CGGCGCGCGG  
 9701 GTGCGGGGCT GGCCGTCGGC AGCCATGCCG GCGAGCGCGG GCTCGGCGCC  
 9751 TTCGGGCGGA CGCTGGGATG CGCGCAGCAG TACGAAACCG ATGGTGCCCA  
 9801 CCGCGGCGGC CAGGGCGAAG AAGACCAGCG TGGGCATGCC GGGAAATGAGG  
 9851 CCCAGGCCTG CCGAGATCGC GCCGGCAATG ACCAGGGCGC GAGGCTGCGC  
 9901 CAGCACTTGT GCGCCGATGT CGGTGCCTAC GTTGAGGGG CCATCCCCGG  
 9951 TCTGCACCCG CGTCACGATG ATTCCGGCGC AGATGGCGAT GAACAGCGCC  
 10001 GGGATCTGCG CGATGAGCCC GTCGCCTATG GTCAGGATGG CATATGTCTG  
 10051 CACGGCCTCG CCGGCGCTCA GGCCGCGCTG CAGCACGCCG ACCAGCATGC  
 10101 CGCCAAGCAG GTTGACGGCA ACGATGATCA GGCCGGCGAT GGCATCGCCC  
 10151 TTGACGAACT TCATCGCGCC GTCCATGGCG CCATACAGTT GGCTTTCCTT  
 10201 CTCGACCGTA CGGCGTCGGC GTCGGGCTTC GTCCATGTCT ATGGTGCCCG  
 10251 CGCGCAAGTC CGCGTCGATG GACATCTGCT TGCCGGGCAT GGCGTCCAgC  
 10301 GAGAAGCGCG CGGCGACTTC GGCCACCCGC TCCGCGCCTT TGGTGATGAC  
 10351 CACGAACTGC ACGATCGTGA GGATGAGGAA AACCACCAgG CCGACGATCA  
 10401 GGTGCGCGCC CACCACGAAG TTGCCGAAGG TCTCGATGAT GTGGCCGGCA  
 10451 TCGCCTTGCA GCAGGATCAG CCGCGTGGTC GCGATGGAGA TGCCAGCCG  
 10501 GAACAGCGTG GTGACCAGCA GGACCGAAGG GAACGAGGAA AACGCCAGGG  
 10551 GCGAAGGCAg GTACATCGCG ACCATCAgCA GGA CTGCCGA CAGCGTCATG  
 10601 TTCGCACCGA TCAGCACGTC GACCAGCGTT GTGGGCAACG GCAGGATCAT

**Fig. 5** 7/23

**THIS PAGE BLANK (USPTO)**

10651 CATGAAGACG ATCGCCACGA TGAGCACGGC CAGTACGATG TCGTTGCGGC  
10701 TGGTGGCCAG CGCCACCGCG CGTTGCAGGC GGCGAATGGA TTTCTTGCTC  
10751 GTCATGGCGG GGTGCGGCTC GCGCAAGACG CCGCCCGTAG CGCCATATAG  
10801 TCACGCATGG CCTGGCGGGC GTCGTCGGGT CGGTTTCAGCG CCTGCATGGC  
10851 CTGCGCCCGG ACCAGATGGC CGGCGGCATC GGGTGTGGCG CGCAGTGCGC  
10901 GCTTGTCCAG CGTGACCAAG GCCATGCGCG GTTCGCCCTG GTGCAGATAG  
10951 CCCAGCGCCA GGGCCAGCAG GGAAGTGGCTG TCGATCGCGT CCAGGGCATC  
11001 CAGGGCCGCC AGCAGGGCAA CCGTCTTGCT CCATTGGCGC TGCAACTGGT  
11051 AGTGGTGCGC AAGCAATTGC AGAAGTTCGC GTACCTGAAG GCTGGGCGAG  
11101 GGTAGGGTAT GCGGCATCAT CCCTGGTAGA GCGCGCTGCG ATACATGGCA  
11151 GCCAGGTCGC GCAGCTTTCC GGCCTCGTTG AGTACGTGCA AGGCGCGGCT  
11201 TAAGGCTGGC GCGGTGTTGC CGGCATGCAG TTCCATGGCC CGGGACAGTT  
11251 CGTCGCGCGC CTGGGCCAGG GCGCGCTCGA ACTGCGCGGG CTGGAATAAT  
11301 GCGCCATCCG TCAGGCGAGG CCGCGCCGAC TCGTCGAGCA TCGCGCTCAG  
11351 GTCCGGGCGT ACGAGCAGGG CTTTCAGATG ATCGACCGCG CCGGTGGCGG  
11401 GCGGTTTCAG CCAGCGCTCG GGTGGCAGGG TGGGGGCGGG CTCGCAGCGG  
11451 GGACCGCGCA CGATGTGGTC GACGCCGCGC TCCAGGCCGA GGTGCATGGC  
11501 ATGCATGCCG GGTACGGCGC TGGACATGGC GTCACGCCTC CAGGCGGTGC  
11551 AGCCAGTTCG TCAGGCCCAG CACCGCCAGG CGCAACGCCG CTGCGTCGAG  
11601 CGAGCGTTCG GGCAGTCGTG TCTGCGCGAC GATCCAGTCC TCGCTGCCCT  
11651 CCGACCAGAG TGACGTCTGG ATGGATGCGG CGCTTCCGCG CTGCCCATGG  
11701 GCCCGTTTCC ATGCCGCCAG CAACACGGAG GCGGCGTCGC GCTCGACCCG  
11751 CTGGGCCAGG TGGACCAGGG CCGCGCCGGC GACGCATTCG ACGCCCAGGC  
11801 GCGCCCCGTT GGACAGCGCC AGCGACGCCG ATCCGACGG CCCGAATGCC  
11851 AGGCCCTCGA TGCCGATATC CTGGCCGAAC TGATGCAGCG CCCTATCGGC  
11901 AGTATTCATG CGTTCTCCAT TGCTATCGCG TTGTCCAGCG CGTCCTGCGC  
11951 GGCCGCCAGG ACGGTGGCGC GCACGTCCAT GTCGGCGTAG ATCTGCGTGG  
12001 GCAGGTCTTT GAGAATCTGG CGTACGCCAC CGAGGAATGC GATGCGCTCG  
12051 GAGAGGGCAT TCGCACCGTG GCGCTCGGCC AGCTTCTCGA AGCGCGCGGG  
12101 CGCAATCCAT TTGTCCCTCG TGATTCCCAC AAGATCGCGC ATCAGGCCCT  
12151 GGGCGTCCGC AACTCCTGC GAGCCTGCGT TGCCCAACCG TTGTTTCAGG

**Fig. 5** 8/23

**THIS PAGE BLANK (USPTO)**



12201	GCATTGCATT	CCTCCAGTAC	CGTGGCGGCC	ACCTCGACTT	GATAGAGATC
12251	GCTCGCCAAC	ACTTGCAGCC	TGACGCCGTC	CGTCGACGGT	GTCGCCGCGG
12301	CCAGGTCGTG	TCCCAGCGCC	TGAATCAGCG	CGCCCAGCGC	GCCGTGGATG
12351	TCGTCTTTCC	CATAGCGTTC	CAGCACCAGG	TCCAGCGTGC	GCGCCAACGA
12401	CAGCTGGCCC	AGGGCGATGT	CGCGGTACGC	GTGCTGGAAG	CCGGCCAGCT
12451	CGTCAGCGGA	ACGCGCGAAT	GCGCCGGCCG	TGGGCAGGGT	GTTGATGCCG
12501	GCGCGGATTT	CGGGGCCATG	GGCGAGCTCC	AGGTCGGCCA	ATGCATCGCG
12551	CAGGGCTTCG	AGCGCGTGCG	GCGCGGCGTC	CTCGTGCTCG	CCGCGCTGCA
12601	GCGCGTGCTG	CAGCGCGAGG	TATTGCTGCG	TGACACCGGG	AAACGCTTGC
12651	GCGGCCAGCT	GCATGGGGGC	GCCCCGGCCG	CGCAGCAGCT	CGGCGGTCAG
12701	GGCTTCCAGT	TTTGCCTGCG	CGTCGGGGTC	GTGGGTGTGG	GAAAACAGTT
12751	CCGCAAGCTG	CGCCGCGTCC	AGCCAGAGCA	TCGGACGTTT	GGCCGTGACC
12801	TTGCGTTCGG	AGTGATGCTT	TTCCTCGGCA	GCCTGCGCCA	TGTGCAGGCT
12851	GAGCTCCTCG	GCCGCGTCCG	CCAGCGATAT	GCCGGTGGGC	GCCGGTGCGA
12901	TGCGCTGGCC	TTGCAGCCAG	CCGGAGGAGG	TGTTGGCCGA	GGCGTCGTGG
12951	CGCCCCTGCA	TGGCGGCGTG	GAAGGGATTG	GGGGCGGCAT	CGATACGAGT
13001	CATGGGGAGT	CCTCGGAGAA	GGAACCATTT	GCCTACTGGT	GCAGTGAGTG
13051	TCGCGCGCGC	GGTCATGGTT	CCCGGAAACG	GGCGCGATAT	TGGGCAATTC
13101	GCAGCCTGGA	ACTTGCCGCG	GGCGCAGGGT	TACTCAGCAT	GCGTCTTTCA
13151	ACTCGAAGGA	GCTCTCATGA	GCATTGATCT	CGGAGTTTCA	CTCACGTCGC
13201	AGGCCGGCGG	CCTGCAAGGC	ATCGACCTCA	AGAGCATGGA	TATCCAGACT
13251	CTCATGGTGT	ATGTGCAGGG	TCGTGCGGCC	GAACCTCTCA	CGGCTCAAAT
13301	GCAGACCCAG	GCCGAAGTGG	TGCAGAAGGC	CAATGAACGC	ATGGCGCAGC
13351	TCAACGAGGT	CCTGTCCGCG	CTGTCCCGGG	CCAAGGCCGA	GTTTCCGCCC
13401	AATCCGAAGC	CGGGCGACAC	CATCCCGGGC	TGGGACAGCC	AGAAGATCAG
13451	CCGGATCGAG	GTTCTCTCTA	ATGATGCGCT	GCGTGCCGCC	GGCCTGACGG
13501	GCATGTTCTG	AGCGCGCGAT	GGCCGGGTGA	CCGGCCCCGA	CGGCCGGGGT
13551	ACGCAGGTCG	TGAACGGCAC	GGGCGTCATG	GCCGGTTCCA	CGACCTATAA
13601	GGAACCTGAA	AGTGCCTACA	CCACCGTAAA	GGGGATGCTG	GATACGGCGT
13651	CCAATACGCA	ACAGATGGAC	ATGATCAGGC	TGCAGGCCGC	CAGCAACAAG

Fig. 5 9/23

**THIS PAGE BLANK (USPTO)**

13701 CGCAACGAGG CTTTCGAGGT CATGACCAAC ACCGAGAAGC GGCGCAGCGA  
13751 CTTGAACAGC TCCATCACCA GCAACATGCG CTAAGCGCTG CACAAGGAGT  
13801 ATTCCATGCA GGAGCAAGGC ATCCAATCCA TCATGCGCGC CGCGGAAGAG  
13851 CTGGTCGAGC AGACCCGCCA GGCGTTGTAC AGCGTCGACG AGATCTACGC  
13901 CCACGTTGGC GTCGACCCCG CTCGCCTGCG CAATCTGGCG GTCGAGCAGG  
13951 CCAGGATAGA GGCCGAGGCC CAGGCGGCGT TCCGTGATGA CCTCGCGGAC  
14001 ATCGAGCGCG AGGCGGCGCG CGTCAAGGCG GCCTGCACCG ATGCGCCGCA  
14051 GGCCCGCAGG GTGCTTCACA ACCACGTCTG AGCGCGGAGG CCTTCCATGC  
14101 CAAAGTCAGC CGACCAGGGC GGCTCCCCGG CGTCAGCTTC GCATGAGGCG  
14151 TTGCGCCATA TTCTCGACGC AGGCGCTTCG ATGGGGGGCT TGCAGGGGTT  
14201 GGACGAGGCG CAGCAGCAGG CGTTGTACGC GATCGGTCAT GGCGCCTACG  
14251 AACAGGGGCG CTATGCCGAC GCGTTGAAAA TGTTCCTGCCT GCTGGTCGCG  
14301 TGCGATCCGC TGGAAGCCCG TTATCTGCTG GCCCTGGGCG CCGCGGCCCCA  
14351 GGAGCTGGGG CTGTACGAGC ATGCCTTGCA GCAATACGCG GCCGCGGCGG  
14401 CTTTGCAGTT GGAATCCCCC AGGCCCTGTG TGCATGGCGC CGAGTGCCTG  
14451 TATGCGTTGG GTCGTCGCCG CGACGCCCTG GATACGCTCG ACATGGTGCT  
14501 TGAGTTGTGC GGCTCGCCGG AGCGTGCGGC CCTGCGCGAA CGGGCCGAGT  
14551 TGCTGCGCAG GAGCTATGCA CGTGCCGACT GAAACGGCGC CATGTCCGCC  
14601 GTCAAGATTT CAATTCGAGG AGGTTTCGATA TGTCTGTTT TCCGACTTCG  
14651 CCCGGCTCTT TCGGGGCCGG CCCTGTCTTT GACTCCGAAT TGCAGGCCCC  
14701 GGCCCCGTCG GCGCAGCGTC GCGGCGGTGC GGCGCCTGTG CCGCCGCCCCG  
14751 TCGATCGGCG CGGCGTCGAG CCGGGAGATC CCACGCTGGG CATGCTGCCC  
14801 GCGCCAGATT TGCTCGCGGG GGGCGCCGTC AGCCGCACCC GCGCGGCGCT  
14851 CGACGATCTG GACGCCGCAC GGCTCGGTGA AGACATCTAC GCCTTGATGG  
14901 CCGTGTTGCA ACAGGCCAGT CAGCAGATGC GGGACGCCGC CCGTATCGCT  
14951 CGTGATGCCG AGGCTACGCG GCAAACGCAG GCTCTCGGCG ATGCGGCCAG  
15001 CCAGATGCGC CAGGCGGCGA GCGAGCGCAT GGCCGGAGCG ATCGTGCGG  
15051 GCGCCATGCA GATAGCGGGT GGTTCGTGTC AGCTGGGGGC GGGCCTGGCA  
15101 GCGGGTTTGC AGGCCATGGG TGGCGCTGCT GCGCAAGCCA AGGGCGCCGC  
15151 ATTCTCCGAG CAGGCCTCGA CAAGCCGCAA GGTGGCGGCC GGCTTGCACG  
15201 ATGCCCCCGA GCTGCAGGCA ACGGTGCAGG CCCGCGCAAC CCAGCTCGAA

Fig. 5 10/23

**THIS PAGE BLANK (USPTO)**

15251 GCGCAAGCGG CCTCGTTTGG TGGCGACGCG GCTCGTTCGT CGGCAAAGTC  
15301 GCAGCGCGTA TCGAGCGTTG CCCAGGCCGG CGCCGCAGCG GCCGGCGGTA  
15351 TCGGCGGCCT GACCAGCGCC GCCCAGGAAC GCCGCGCCGC CGAGCACGAG  
15401 GCCAGGCGCG CGGAGCTGGA CGTCGAAGCG AAGGTGCATG AAACGGCCTC  
15451 GCGGCGGGCC GACGAAGCCA TGCAGCAGAT GCTCGACATC ATCCGCGGCA  
15501 TCAGGGAAAA GCTGGCCGGG ATGGAGCAGT CCCGCAGCGA GACCGCCCGT  
15551 AGCGTGGCCC GCAATATCTG AGTGTCCGGC TCCAACCTTC AATCTTGAGG  
15601 ATGACCGTCA TGAGTACGAC CATATCCACA GCCCCGAGCG GCGCCGCGCT  
15651 TGCGCCGTCT CGCATAGATA TGGGGCGCC GGAGCCCGGG AGTGCCGGCG  
15701 AAGGCGCCGG TATCCTGGCG CCGGTGACGA CGCTGGCTCT GGCGGCGGGC  
15751 CGGCCGGCTT TGCCAGCGTC ACCGTCGCTG CGCACCGCGC CCGTCCTGGA  
15801 TCCGCCAGTG CGCGATCTCA GCCCCGCCGA CTTGGCCGAC CTGCTGCGCG  
15851 TCTTGCGATC CAGGGCGGTG GACGGGCAGT TGGCCACGGC GCGCGAGAAC  
15901 CTGCAGGATG CGCAAGTCAA GGCGAAGCAG AACACCCAGG CCCAGCTCGA  
15951 CAAGCTGGAC GCATGGTTTC GGAAGGCTGA GGACGCCGAG AGCAAGGGCT  
16001 GGCTGAGCAA GGTGTTTCGGC TGGATCGGGA AGGTGCTGGC GGTCTGGCA  
16051 TCGGCCCTGG CTGTGGGCTT TGCTGCCGTC GCCAGCGTGG TCACCGGCGC  
16101 GGCGGCCACG CCCATGCTGG TGCTCAGCGG CATGGCATTG GTCAGCGCCG  
16151 TGACATCGCT GGCCGACCAG ATATCGCGAG AGGCGGGAGG GCCGCCTATC  
16201 AGCCTGGGCG GGTTCCTCTC CGGGCTGGCC GGACGTCTGC TGACAGCGTT  
16251 GGGGGTGGAT CAGTCGCAGG CCGACCAAAT TGCCAAGATC GTCGCCGGCC  
16301 TGGCCGTGCC CGCCGTCTTG CTGATCGAAC CCCAGATGCT GGGCGAAATG  
16351 GCCGAAGGCG TGGCCAGGCT GGCGGGCGCC GGCGATGCCA CCGCGGGATA  
16401 CATAGCCATG GCGATGTCCA TCGTGGCGGC GATCGCGGTC GCCGCGATCA  
16451 ATGCCGCCGG TACGGCCGGC GCGGGCAGCG CCTCGGCGAT CAGGGGTGCC  
16501 TGGGATCGGG CCGCCGCGGT AGCCACCCAG GTCCTTCAGG GGGGTACGGC  
16551 AGTGGCGCAA GGCGGCGTCG GCGTGTCGAT GGCAGTGCGT CGCAAACAGG  
16601 CCGATCTCCT GGTGCGCGAC AAGGCGGATC TGGCGGCGAG CCTGACAAAA  
16651 CTGCGGGCGG CCATGGAGCG TGAGGCGGAC GATATCAAGA AGATCCTGGC  
16701 TCAATTCGAC GCGGCCTATC ACATGATCGC GCAGATGATC AGCGACATGG

**THIS PAGE BLANK (USPTO)**

16751 CGAGCACGCA CAGCCAGGTC AGCGCCAACC TCGGACGGCG CCAGGCGGTG  
 16801 TAGCGCCGGG CGCTCAAGGA ATTTTCATGA CTGTTCATGA CGACGCGGCC  
 16851 GCGGCGCTGC GCGCCCGGCT GGATGCGTTG CCGGGCAGCC GGCGCCTGAC  
 16901 AGCCGAGCAA TTGGAAGTGA TTTACGCGAT GGCCTATGCG CACGTCGCCA  
 16951 GGTGCGAGTA CGGCAAGGCG CTGCCCATTT TCGCCTTCCT CGCGCAGTAC  
 17001 GGCCCCACGC GCAAGCACTA CTGGGCCGGC CTGGCGCTAT GCCTGCAGAA  
 17051 GACCGACCGT CCCGACGAGG CGCGCAATAT CTATGCGTTG ATCCTCACGC  
 17101 TCTATCCAGA TTCCGCGGAT GCCGTGTTGC GCACGGCCGA GTGCGAGCTG  
 17151 GCGTTGGGTG AGAACGAACG GGCGCAGGCG GCCCTGTTCT GCGCAATCGC  
 17201 CATCGATGCA GAAAGTGGGC AGCCAGGTCC GGTCTCGCAC CGTGCGCGCG  
 17251 CTTTGCTCGA TCTTATTTCA GTTTCACATC CGGAGTAACT CCATGCACTC  
 17301 AGACTCAGGT TCAGATTGAG GCTCAGACTC AGGCTCaGGC TCACCCATGG  
 17351 TCTCGTCGAT ACATCCATCG GAACCGATAC AGCCGATGGA GCATGTGCTC  
 17401 GAGGAGGCCG ACGCCCGCCT GCTTACCGAA GTGGGTTTTT TGGCGGGCGC  
 17451 CGTCAGCGAT CTGACGCGCG CGGACGCCAT TTTCAATGCA TTGCAACGTG  
 17501 TACGGCCGGG CCGGACGCAT CCCTGCATCG GCCTGGCGGT CGCCCGCATG  
 17551 AACGCCGGGC TGCCCGACGA AGCCGCCGAG ATCCTGGCGA ATTTCCAGCC  
 17601 GGCACAGCCG GAGGACCGCT CGGAACTGGA CGCCTGGTGC GGGTTCGCTC  
 17651 TGTGCTGGC TGGCCGCTCG GACGAGGCGC GCCGCATGCT GCAGCGAGCC  
 17701 ATCGATGCGG GTGGCGAGGC GGCAAGGCTG GCGCAGGTG TGTGACAG  
 17751 CGGACCCGCC ATGATGCGGC CCGCGCCGTT GCAGTCCGAG CCATTACCTG  
 17801 GAGCTCCTGG ATGAATTTGG ATCTGACGGC GATCAACGCC GTGCAGGAAC  
 17851 GGCTGCTCGC TGGATCATTC GACATGCCGC GATCTCCCGC GATGGCGGAT  
 17901 CAGGCGCGCT TTGAATTGGC GCTGGGCGAG ATGCCCGGCG CATCGGCCCC  
 17951 GAACGGGGCG ATCGCCCTGG CGCCGGTCGC GCTCGACGAG CCGCTGGGCC  
 18001 GTCGATTCT TGGACAGTTG CGCGGCGGCC TGGCCGATGT GGCAGGAAAA  
 18051 TGGCGGGCGG TGCAGACGGG CTTGGCCGAG GTGAGCCAGG CGCCTACCGT  
 18101 GGTGGGTATG CTCGATCTGC AGGCCAGGTT GCTACAGGCA TCCGTGGAGT  
 18151 ACGAGTTGGT GGGCAAGGCA ATAGGGCGCG CCACCCAAAA CGTCGATACG  
 18201 CTGGCGAGAA TGTCATGAAC GCCATCGGGG CGATCCAACG GTATCGGCGC  
 18251 GGCGCGGGAT GGGCGGCCCT GGTGCTCGCC CTGGCGCTGC TGGCCGGCTG

**THIS PAGE BLANK (USPTO)**



18301 CCGTGCCCGC GTCGAGCTGT TGGGCGCGGC GCCCGAGAAC GAAGCCAACG  
 18351 AAGTATTGGC GGCGCTGCTC GAGGCAGGCA TCGCTGCGCA GAAGCAGTCC  
 18401 GGCAAGGCCG GCTACGCGGT TTCGGTGCCG GCCGAGGCGG TGGCCCGGTC  
 18451 GCTGGAGATC CTGCGCGCAA GCGGCCTGCC CCGCGAGCAG TTCGACGGAA  
 18501 TGGGACGCAT ATTCCGCAAG GAAGGCCTGG TTTCATCTCC GCTCGAAGAG  
 18551 CGCGCCCGCT ACATTTATGC GCTGTCTCAG GAATTGGCCG ACACCCTGTC  
 18601 GCAGATCGAC GGCGTGCTCA GCGCCCGCGT GCACGTGGTG CTCCCCGAAC  
 18651 GCGGCGCGGT CGGCGAGCCG GCCACCCCTT CGACGGCAGG GGTGTTTCTC  
 18701 AAGTACCGCG ACGGACAGAG CCTCGACGCG CTCGTGCCCC AGATCCGCAA  
 18751 GCTGGTCACG CATGCCATCC CGGGCCTGGC CGAGGACCGT GTATCGGTTG  
 18801 CCCTGGTGGT GGCCAGCCC GTTCAGGCCG CACCCGCGCC GGTCGCGTGG  
 18851 CGCCGCGTGC TTGGCGTACA GGTCGCGGAC GGATCGGTCC TGAGATTTTC  
 18901 GCTGTTGCTG CTGTTGTTGC CGGTGCTGTG CCTGATAGTG GCGGGGgCCG  
 18951 CGCTCTACGT CTGGCGCACG CGCTGGTCCC GCGGCGAAGG GCGCGGCGGG  
 19001 GCTGGCGCCG GCGCCACGGA AGGAGCCGGG CATGACTGAG GCGAGCGTGC  
 19051 TGCTTTCCGA GCGGCTCATG ATATTCAATC TCCTGCCCAG CCTGACCCTG  
 19101 CATGCCAGTC GCCACGACGA GATGTTTCCA GCCGATTGGG TGCGCGCGTT  
 19151 GTGCAATGCC GACGCGGCGT TGGCCAACGC GTGGCATCGC CATTGGTCGC  
 19201 GCTGGATCTT GTGCGAATTG GGCCTGCTGA ACCAGCCGGT CCTGAGCCTC  
 19251 GATCCGCCGC AGTTGAAGGT CGCGCTATTG TCCACGGACG CCTTGCGGAC  
 19301 CTGCGCCGCC CATGCGGGAG CGCTGCTGTG CGCGCCGCGC CTGCGACGCG  
 19351 CGATAGACGG CGCCGAGGTC CGTACCTTGC ATGCCGCGCT CGGGCGCGAT  
 19401 GTGATGAATT TCGCCGTGTC TTCCGCGGCG CGGGCCCTGC ATGACGGGCT  
 19451 CGCCGCCAGT TCGGACTGGA CCCTGGCCGC CACGGTCCAG GCGGCGCAGA  
 19501 AACTGGGcTG GGCCCTGCTG CGCGACGCCG TGCAGGGCGC CGCCGACGAG  
 19551 ATAGCGCTGC GTTGCGCGCT GAAGTTGCCG CGCGACCTTG ATCCCGCGCC  
 19601 CGTCCTGCCG CCCGAGGCGG CGCTTGCGCT GGTGCTGTCC ATGCTCGAAA  
 19651 TCCTGGATGC AGAATGGCTT TCCTCGTTCC CCGCCCAAGC CTGATCCAGG  
 19701 CCGTACGGCC CGGCCGTGCG GATCCCGCGA CCGACGTCTT GCGCGCTGAA  
 19751 GACTACGCCG AGCTGCTCAG CGCCGCGCAG ATCGTTGCCC AGGCACACCG

**THIS PAGE BLANK (USPTO)**

19801 GCGGGCCGGC GAAATCGTGG CCGAGGCGCG AGAGGAGTTC GAGCGCGAGC  
 19851 GCAGGCGAGG CTATGAGGAG GGGCGCCGCG AAGCGCTTAC GGATCAGGCG  
 19901 GAGAAGATGA TAGAAACCGT AAGCCGCACG ATCGACTACT TCGCGGGTAT  
 19951 CGAGAACGAG ATGATCGAAC TGGTCATGAG TCGCGTCCGC AAGATCGTCG  
 20001 ACGGTTACGA CGACCGCGAG CGCACCCTGA TCGCCGTGCG CAACGCATTG  
 20051 GCGGTCGTGC GCAATCAGCG CCAGATGACC TTGCGCCTGC ACCCAGACGA  
 20101 GGTGGATGTG CTCCGGGAAG GCATGAACCA GCTTCTGGCG GCCTATCCGG  
 20151 GCGTGGGCTA CCTGGACCTG CTGCCCCGACG CCAGGCTGGC GCCGGGAGCC  
 20201 TGCATACTGG AGAGCGAGAT AGGCATGGTC GAGGCCAGCC TCGAGGACCA  
 20251 GCTGTGCGCC TTGCGGGCGG CCTTCGAACG TACATTCTGGC CgGCGCGGAT  
 20301 AGGGGCATGC GTCaGTACCa CTACATCACG GA<sub>g</sub>ATGATGC GGGTGGCCCT  
 20351 GCAGGATCTG TCCACGCTGC GGATAAAGGG CCGGGTGGTG CAAGTGGTGG  
 20401 GAACGATCAT CAAGGCCGtC GTTCCGATGG TCAAGATCGG CGAAGTGTGC  
 20451 CTGCTGCGCA ATCCCGGCGA GGA<sub>g</sub>CTTCGAG ATGCACGGCG AAGTGGTGGG  
 20501 CTTTGTCCGC GACGCCGCCT TGCTCACGCC TATCGGCGAC ATGTACGGGA  
 20551 TTTCTCTGGC GACCGAGGTG ATACCGACCG GACGCACGCA TATGGTCCCC  
 20601 GTCGGTCCGG GCTTGCTGGG ACGCGTGCTG GACGGGCTGG GACGTCCGCT  
 20651 GGACGCCGCC GAGTCAGGGC CGTGCAATGC CCACAAGTTC TATCCGGTCT  
 20701 TCGCCGATGC GCCAGACCCG CTGACGCGTC GCATCATCCA TGCTCCGCTG  
 20751 GAGCTGGGGG TGCGCGTACT GGACGGTTTG CTTACATGCG GGGAAGGCCA  
 20801 GCGTCTGGGA ATTTTCGCAG CCGCCGGCGG CGGCAAGTCG ACCCTGCTGG  
 20851 GCATGCTGGT CAAGGGCGCC GCGGTCGACG TGACGGTGGT GGCGCTGATC  
 20901 GGCGAGCGTG GGCGGGAAGT TCGCGAGTTC CTTGAGCACG AACTCGGTCC  
 20951 GGAGGGCAGA CGCAAGAGCG TGATCGTCTG CGCGACCAGC GACAAGTCCT  
 21001 CGATGGAGCG TGCCAAGGCG GCGTACGTCG CAACCGCCAT CGCCGAATAC  
 21051 TTCCGCGATC AAGGGCAGCG TGTA<sub>g</sub>CTTTT CTGATGGACT CGGTACCCG  
 21101 CTTTGC<sub>g</sub>CGA GCCCAGCGTG AAATCGGCTT GGCGGCAGGC GAGCCGCCGA  
 21151 CGCGGCGCGG CTATCCACCG TCGGTGTTCG CCACCTTGCC CAAACTGATG  
 21201 GAGCGCGCCG GCATGAACCA GACGGGTTCG ATCACGGCGC TGTATACGGT  
 21251 GCTGGTCTGAG GGGGACGACA TGAACGAACC GGTGGCCGAC GAGACGCGTT  
 21301 CGATACTGGA CGGCCACATC GTGCTCTCGC GCAAGCTGGG AGCGGCGAAT

Fig. 5 14/23

**THIS PAGE BLANK (USPTO)**

21351 CACTATCCTG CCGTCGACGT GCTGGCCTCG GCCAGCCGGG TCATGAATGC  
 21401 CGTGGTGTCTG CCGCGTCACA AGTACCTGGC CGGACGTATG CGCGAACTGA  
 21451 TGGCCAAGTA CCAGGATGTC GAGCTGTTGG TGAAAATCGG CGAGTACAAG  
 21501 CAGGGCGCCG ATGCGTCGAC CGATGAGGCG ATACAGAAGA TCGGACAGAT  
 21551 CAATGCGTTT CTCAGACAAC TAACCGACGA ACGCGAAGCA TTCGAGGATA  
 21601 CCGTACTGCG CATGGCTGAA ATCATCGGAC CCGAATCCTA ATGGACCTGG  
 21651 AAAGCCTGCT TGCCATCAAG CATTTCGCG CCGACCAAGC CCAGCTTGCG  
 21701 CTGAAACGCC AACAGCAGGC CTGCGCGGTT GCTGCCGCGG CGCAGCGTCA  
 21751 GGGCGAAGGC CGCCTCGACG ATTGTCGCCT GTGGGCCGGA CAGCTCGAAA  
 21801 ACCGTCTATA TGCCGAGCTG TGCCGGCGCA TCGTCAAGAC ACGCGACATC  
 21851 GACGAGGTGC TGCAACGAGT GGGCCACGCC CGCGACCGCC AGGCCAGCCT  
 21901 GGGCGTGCAG CTCGACGACG CCGTGCGCCG TCACGAACAT GAAATCCAGC  
 21951 TGCTCGCGCA GCAGCGCGAG CAGCACCGGG AGTGCTTCCA GGCGCAGCAA  
 22001 CGGATCGCCG AGTTGGTGCG CCTGCAGCAG GTCGAGGCGG CGGCCTTGCG  
 22051 CGAGAGCCAG GAAGATCGCG AAATTCAGGA AGCCATCGAA TTGTCGGCGC  
 22101 GTGGGCGCGA CGATGCATCG CGAGCCGGCG ACGGCCTGGC GCGGCTATGA  
 22151 ACCAGCCAGA CGGGCTGGGT TCGCCCATGG CCGGCGGCGG GCAGCGCATG  
 22201 GGGCTGGCGC GCACGCCGTA TGCGCGTCAG CCGGATCGGG ATGCGCAGCG  
 22251 TGCCTTCGAG CGGGAAATGG AACAGGAGAA AGCGAAGGAA GAACTGCCCC  
 22301 GGCCGCAACG CCTGGCGCCG GGTCCGGCCT GCGTCGGCTG GCTGGCGTCG  
 22351 ATGGAACCTG CCGCCGGCCG TCCACCGGCC AGTCTGGCCC AGGCGCTGGC  
 22401 AAGCGTGGCT GCGGGGCTGG CGGTAGGCGA CGTGCTGGAG GGGTATCGCG  
 22451 AAGCCCGTAT CGTTGTGGAC GATACGCTGC TACCCGACAC CACCTTGTCG  
 22501 GTACGGGAGG ACGGCGGCTG GATCGTGGTG GCTTTCGCAT GCCGACAACG  
 22551 GGACGCTTGC GAGCGCCTGC ACGCGTGC GCACCGGTTG GCCATGGAGC  
 22601 TCGCGCTGGA GCTGGCGCGC GACGTCGAGG TTGCGGTGGC ATGCGACGGC  
 22651 GAGCCGCACG AGCGGGTGGC GCGCGCGCAG CGGCCGTGGC GATGAATCGA  
 22701 GTGGCCGGCG GGGCGGCGGC GCAGGCCGCT GGCATGGTGG ATCTCGCGGT  
 22751 TCCGCGGTTG AGCGCCGGCG AGGCCCATGC CCTGTCGAGG ATTGCATGCC  
 22801 ATGGCGCGCG ATTTCACGTT CGGCTTGGCG AGCCGGCCGT GCGCTGGCAC

Fig. 5 15/23

**THIS PAGE BLANK (USPTO)**

22851 TGCGCCCTGA CGCCTTGCGT GCACGGCGAC CTTGCCGATG GCGAGATGGA  
 22901 AAGCCTGCAA CTGCAATGGG CCGGGACGTA CATCGGCCTG ACGGTTCCGC  
 22951 GCGCGGCCGC GCGGGGATGG CTGGCGGCGC GCCTGCCCCG GTTTTCCGGC  
 23001 GTGGAGTTGC CGGAACCCAT TGCGGCGGCG GCCCTGGAGG CAATGCTGGA  
 23051 GGAGGTCTGT CGAGGCCTGG CCGGACTCGA CCAGCAAGGC CCGGTCCGCG  
 23101 TGGCGCGGCA AGGCGGGACG CCACCGGTCC AGCCGCATCG CTGGACCCTG  
 23151 ACGGTACGGG CGCCTGACGG TGGCGTCTGG CGCGCGGTAC TGGCGTGCGA  
 23201 CGCATGGGCC TTGCAAGCGG TCGCGGCGGC GCTGGATTCC GTTGCGCGCTG  
 23251 CCGATGGTCG GGTCAATCCG GAGCGCGTGC CGGTCAGGTT GCGTGCCGAT  
 23301 GTCGGCGCGG CGTCCGTGAC CGCAGGCCAG CTGCGGACGC TCGAGCGGG  
 23351 CGACGTCGTG TTGCTCGCGC AGTACCGGGT GAGCGATGCC GCAGAACTAT  
 23401 GGTTGTGCGC CGGACCCAGC GCGATCCGGG TACGGGCCGA GCATGCGTCT  
 23451 TTTCGTGTAA CTCAAGGTTG GACTCCCATC ATGACGGAAC CCGCGACACC  
 23501 TGACCCTGGC GAAACCCCGG CACAGGCCGA CGCGACGCTC GATACCGATC  
 23551 AGATACCCGT GCGCCTGACG TTCGACCTGG GCGAGCGCGA GTTCACGCTT  
 23601 GCGCAGCTGC GCAGCCTGCA TCCGGGCTGC ACGTTCGACC TCGAGCGGCC  
 23651 CATCGCCGAC GGGCCGGTCA TGGTGCGGGC CAATGGCCTG TTGCTGGGCA  
 23701 GCGGCCGGCT GGTCGACATC GACGGCCGCA TCGGCGTGGT ATTGCAGTCG  
 23751 GTCAGGCCTG GACTCGCATG AGCGATACCG ACCCCTTCAG CCTGGCCCTG  
 23801 TTTCTGGCGC TGCTGGCGCT GGTACCGCTC ATCGTCGTCA TGACCACGTC  
 23851 GTTCCTGAAG ATCGCCGTCG TGCTTGCCCTT GGTGCGCAAC GCCCTGGGAG  
 23901 TGCAACAGGT ACCGCCCAAC ATGGCCCTGT ACGGGCTGGC GCTTATTCTT  
 23951 TCCGCGTATG TGATGGCGCC GGTCGTTCAC AGGATAGGCA CCGAGGTCCA  
 24001 GGCCTTGACC GCGCAAGCCG GGGAGTCCGG CACCGCCGCG CCGATGGCGC  
 24051 TGGACGCCGT GCTTGGCGTG GCCGAGCGAG GCGTGGGGCC GCTGCGGGCC  
 24101 TTCATGTTGC GCAACAGCCA GCCGGCCCAG CGTGATTCTT TCCTGCGCAC  
 24151 AGCGCGTCAT CTCTGGGGCG AGGAGGCATC GCGGGACCTG TCGGAAGACA  
 24201 ACCTGCTGGT ATTGACGCCC GCATTTCTGG TTTCGGAGCT GACCGCCGCA  
 24251 TTCCAGCTTG GCTTTCTGCT GTACCTGCCG TTCATCATCA TCGACCTCAT  
 24301 CGTATCGAAC ATTCTTCTTG CCATGGGAAT GATGATGGTT TCTCCCGTGA  
 24351 CGATCTCCAT GCCGTTGAAG CTGTTCTGT TCGTCATGGT GGACGGCTGG

Fig. 5 16/23

**THIS PAGE BLANK (USPTO)**



24401 ACGCGCCTGA TCCAGGGCCT GGTGCTTTCC TATCGGTGAC CAGCATGCAA  
 24451 ACCCAAGACC TGGTTTCGTT CATGACACAG GCGTTGTACC TGGTGCTCTG  
 24501 GCTGTGCTG CCGCCCATCG CCGTGGTGGC GATCGTGGGA ACGCTGTTTT  
 24551 CCCTGTTGCA GGCCTTGACG CAGGTGCAGG AGCAGACCCT GTCCTTCGCC  
 24601 GTGAAGCTGA TAGCCGTGTT CGCCACGCTG ATGCTGGCGG CCCGGTGGAT  
 24651 AAGCGCGGAA ATCTATAACT TCACGATTGC GGTGTTTCGAT GCCTTTCATC  
 24701 GGATCCACTG AGCGGCCAAT CGATGCACAC GGAGTTCAAT TTCGTCGAGG  
 24751 CGAAGGTTTT CCTGGGAACG CTGGCCATGA CGCAACCGCG GATACTCACG  
 24801 GCCATGCTCT TTCTGCCGAT GTTCAACCGT CAGTTTCTGC CTGGTCCGCT  
 24851 GCGTTACGCC GTCGGCGCCT GTCTCGGGCT GATCGTGGTT CCCCAGCTGG  
 24901 CGCCGCAGTA TGCCGCGCTG GATATCGACT GGCCCCGGCT GCTGGCGCTG  
 24951 CTGGCCAAGG AGGCGATGGT GGGCATGTTT CTGGGTTGGC TGGCTGCCTT  
 25001 GCCATTCTGG ATCTTCGAGG CCATCGGCTT CGTCATAGAC AACCAACGGG  
 25051 GCGCCAGCCT GGGCGCTATC CTCAACCCCG CCACGGGCAA CGATTCGTCG  
 25101 CCCATGGGCA TTCTCTTCAA TCTGGGATTC ATGGTGTCTT TCCTGACGGC  
 25151 GGGCGGATTC GGGTTGTTCG CCACGATGCT GTATGACAGC TTCGGGTGTG  
 25201 GGAACATCTG GGCGTGGTGG CCGTCCATGC CCGCACAGGG CGCCGTGCGG  
 25251 ATGCTGGACC AGTTCAGTGG CTTTGCCGCG CGTGTCTGCT TGCTGGCCTC  
 25301 GCCGGCCATC GTGGCCATGT TCCTGGCCGA GCTGGGCCTG GCCCTGATCA  
 25351 GCCGCTTCGC GCCTCAACTG CAGGTGTTCT TCCTGGCTCT GCCGGTAAAG  
 25401 AGCGCGCTGG TGCTGTTCGT GCTGGTGCTG TACATGGCAA CGTTGTTCCA  
 25451 GTATGCAGGC GAAATCCTGG GTTCTGTGGG CCGGATCGTG CCGTTCCTGC  
 25501 ATTCAGCGTG GCCCGGCCCA TGAGCGGCGA GAAAACCGAG CGGCCCACCC  
 25551 CGAAGCGCCT GCGCGATTCC CGCGAGAAAAG GCGAGGTCGC ACACAGCCGG  
 25601 GACTTTACCC AGACGGCGCT GATATGCGCC TTGTTTCGGGC ACTTTCTGAT  
 25651 CAATGCCCCG TCCATTCTCG CGTCGCTGCG AGCGCTGATA CTGGCGCCGG  
 25701 CGGCCTTTGC CGACCAGGGG TTCGCCGTCG CATTGGGGCC CGTGCTGACG  
 25751 GAAATCCTCG ATCAGGCCGT CCGCGTGCTC GCTCCGCTGA TTCTCATCGT  
 25801 GCTTGGGGTG GGGATGTTCG CCGAATTCCT GCAGGTAGGC GTCGTGCTGG  
 25851 CGTTTCGAAA GCTCAAGCCT TCGGCGGAGA AACTGAATCC CGCCGGCAAT

**Fig. 5** 17/23

**THIS PAGE BLANK (USPTO)**

25901 TTGAAGAATA TCTTCTCGGC GCGCAACCTG ATGGAGTTCA TCAAGTCGGT  
 25951 ATGCAAGATC CTGTTTCTGG CCGTGTGGT CACGTTGGTG ATACGGGATT  
 26001 CCTTGCAGCC GCTGATGGCC GTTCCCCATA GCGGGCTGGA CGGGTTGCGA  
 26051 ACGGGCGTAG GCCGATTCT GCAGGTCATG GTCTGGAACA TCGGACTGGC  
 26101 GTACGGGGCG ATTTCTGCTGG CGGACCTGGC CTGGCAGCGT TACCAGTATC  
 26151 GCAAAGGCTT GCGGATGAGC AAGGACGAAG TGAAGCAGGA GTACAAGGAG  
 26201 ATGGAAGGCG ATCCCCATAT CAAGCAGCAA CGCAAGCACC TGCACCAGGA  
 26251 GCTGATCATG CATGGCGCGG CGGCCAGGT TCGCCGGGCG ACGGTGCTGG  
 26301 TGACCAATCC GACACACCTG GCCGTGGCCC TGTACTACGC GCGGGGCGAG  
 26351 ACGCCCTTGC CGCGCGTGCT GGCCATGGGG CAGGGAGCCG TGGCCGCTCT  
 26401 CATGGTCGAG GCCGCGCGCG ATGCCGGCGT GCCGGTCATG CAGAACGTCG  
 26451 CGCTGGCCCG CGCCTTGAC GACCAGGCGG AGGTGGACCA ATACATTCCC  
 26501 GCGGAGTTGG TGGAGCCGGT GGCCGCGGTG TTGCGGGCGG TCGCCAGGC  
 26551 ACTCAAGGAG CAGACATGAC AGCAACCATT CATCCCGATA TTGCCGATTA  
 26601 TGCGCGACGC CATGGCCTCG AACCTCGGT CGACGCCGAT GCGGGGCTTG  
 26651 CCGTCCGGAT CGACGGACGG CATCGCGTCA GGTGATCCC CGCCGAAGAC  
 26701 GGCATGCTGG TGTGCGGGC GCGGCTGGCC GAGCTGCCC GATGGGTGGCA  
 26751 GCGCGCGCG GCGCAGTTGC GCCGGGCGGG CCTGCTGGCC AGCGCCATGG  
 26801 CCCCTGCGAC CGATGCGTAC TGCGGCATAG ACCAGGGCGA AACCGCGTTG  
 26851 TATCTGCACC AGCGCGTCGC ACCGGCCGGC AGTGCGCTGG CGGTGGACGA  
 26901 GCGGGTGGGC GAGTTCGTCA ATGCCTTGGC CACTTGAAA AGGGCGATGG  
 26951 CGCAATGGCA ATAGGTCGGC TTGGGTATCT TGTCCGCGGC GCATGGGCCG  
 27001 GGGGTGTCAT GCTGTTGGCG GCCGGTAGCG CCTGGGCGGC GCCGAACTGG  
 27051 CCTTTGGCGC CGTATAGCTA CTACGCGCAG CAGCAGAGCC TGTCCGATGT  
 27101 GCTGCGCGAG TTCGCCGAG GCTTCAGCCT GGCGTTGCAA CAGGGCAAAG  
 27151 GGGTGCAAGG CGTGGTCAAT GGGCGTTTCA ATGCGCGCAC ACCCACGGAG  
 27201 TTCATCGAGC GTCTCAGCGG CATCTATGGG TTCAACTGGT TCGTGCATGC  
 27251 CGGCACGCTG TATGTCAGCC GCACCAGCGA CGTGGTTACC CGCGCGGTGG  
 27301 ATGCAGCCGG CGCTTCGCCG TCGGCGTTGC GCCAGGCCTT GCTGCAACTG  
 27351 GGCATCCTGG ACGAACGCTT CGGATGGGGA GAGCTGCCGG CGCAAGGCGT  
 27401 GGCCATGGTG TCAGGGCCGC CGGCCTATGT CGCGCTGGTC GAGCAGGCCG

**Fig. 5** 18/23

**THIS PAGE BLANK (USPTO)**

27451 TAGCGGCGTT GCCCAAGGGG GCCGGCAATC AGCAGGTGGC GGTGTTTCGC  
 27501 CTCAAGCATG CTTCCGTGAG CGACCGGGTG ATCCGTTATC GAGACCAGCA  
 27551 GGTAGTTACG CCGGGGATGG CCACCATGCT GCGCCAATTG ATCCTGGGGG  
 27601 CGGGGCCGGG CAACGACGCG GCGCTGGCCG CGGTGGCGGC GCCGCTGCGG  
 27651 GAAATCCGC CGGTGTTTCG CGATGCGGCA GCTGACGGGA ACGCGCCGCT  
 27701 CGCTGGCGCA GCCCAGGCAG CCGGCCGGCG CCTGAGCGAG CCCAGCGTGC  
 27751 AGGCCGACAC GCGCCTCAAT GCCTTGATCG TGCAGGATAT TCCCGAACGG  
 27801 ATGCCAATCT ACCGTGCCCT GATCGAGCAG TTGGATGTGC CCAGCACCCCT  
 27851 GATCGAAATA GAGGCCATGA TCGTGGACGT CAATACCGAT CTGGTCAACG  
 27901 AGCTGGGTGT CACCTGGGGG GCGCAGATCG GAACCACCAG CCTGGGCTAT  
 27951 GGCGATCTGG GGCTGCGTCC CGGCAACGGC CTGCCCCTGG ACGGCGCGGC  
 28001 GGCCGACCTG GCGCCCGGAA CCTTGGGGAT CAGTGTCAGT ACCCGGCTGG  
 28051 CGGCGCGCTT GCGTGCCTTG GAGTCGGACG GGCAGGCCAA TATCCTGTCT  
 28101 CAGCCGTCCA TCCTGACCGC CGACAACCTC GGCGCCATGA TAGACCTGTC  
 28151 GGATACCTTC TACATTCGCA CCCTGGGCGA GCGCGTAGCG ACAGTCACGC  
 28201 CTGTACGGT GGGTACGTCG TTGCGTGTGA CGCCGCGCTA TATCGCCGCC  
 28251 AAGGGAGGAC GCCAGGTGGA ATTGGCGATC GATATCGAGG ACGGACGGGT  
 28301 CTTGCAGGAG TATCCCATCG ATGGTCTGCC CCGGGTTCGG AAAAGCAGCA  
 28351 TCAGCACGCT GCGGGTGGTG GGGGACGAGC AGACGCTGCT GATCGGCGGC  
 28401 TACAACAATC GCCGTGACGA AGAGCAGGTC GAGAAAGTGC CGCTGCTGGG  
 28451 AGATATCCCC GGCCTGGGGT TCTTGTCTC GAGCAAGTCC CGGGCGGTAC  
 28501 AGCGCCGCGA GCGGCTGTTC CTGATCCGGC CGCGTGTCGT GGCTATCGAG  
 28551 GGCAAGCCGG TCTTCAGCCC CGTTGCGGGC ACGTCGCAGG TGTTTCATGAG  
 28601 CACGGGTTGG GCGGGGATG GCAGCAGCCT GAGCATTGCA CCCGCGGAGG  
 28651 GCGGGCATAA ACAAGTGCGT CATGATGCCC GGGCGGGCAG GCCGGTCCGG  
 28701 CTGGTGCCGG ATTCATTGCA TGTGGAGTAT GGCGAGGCGG GGGAGGCGTC  
 28751 GCCCTGAGCG TGGCCCGGGC AGGGGGGCTA CGGCACGCTG TCGTAGCCTC  
 28801 GTTCGCGCAG CGCTTCGCGC AAGGCACAGC GGGCGCGGGA AAGTCGGCTG  
 28851 CGAATGGTTC CTACAGGAAC CGAAAGCAGT GCGGCAGCCT CTTCATAGGA  
 28901 GAGTTCTTCC ACGCCGACCA TGAGGATCAC GTCGCGCATG CTTTCGGGGA

Fig. 5 19/23

**THIS PAGE BLANK (USPTO)**

28951 GCTGCTCCAG CGCTTCGCGT AGCAAGCGCA TCGGTTGACG CTGCTCGGTC  
 29001 ACGGCTTCGG GGTGGGGCGC ACTGCATGGC ATGACACCCA GGCTGGCGTC  
 29051 GgAGGTGAAT TCATAACGGC GCTCTGGCGC ACGCGACAAG TGATTGCGGA  
 29101 CCAGATTGAG CGCGATGCCG TACAGCCAGG TGGAAAGCTG GGAGTCGCCA  
 29151 CGGAACGATT GATACGCGCG CGCCGCCTCG GCGAAAGCCT GCTGCGCAAG  
 29201 GTCCTCGACG TCGCTGCTGT GGCCGATGTG CTTGGCGATG AATCTCTGTA  
 29251 ATCGGCCGGC ATGTTTCGGCT ACCAGGCTGC TCAGCAACTG CTGGTCCGAT  
 29301 CGCTCGGTGG CGATATCGGG ATATGCGTTG TCCGGTTTTT CGAGAGAAGC  
 29351 GGGAGACCG GATGACGGGG GAGAAACCAT GCAAAGCGAT ACCAAGTGAA  
 29401 AGGGTGATAA TTCACGTCAC CAAGATACTG ACTGCCGGTT TTATCCGGCA  
 29451 GTTGTTAACT TCCGAAACTA ATGTCCGATC GCGGTCGTA CCGGAGCATT  
 29501 CAGATACAAC GCGCTGAACG TCTTCCGTAA AACTTACGAC GCGACGTATG  
 29551 GGGACTACAC CAGGTGGTGC GCAAGGACCT CGCGCaaCCA TTCTTGCGCC  
 29601 GTATGGGCGG ATACCGGCGC CTTACCTGAT TGTGCGGCGG ATGCGAAGCG  
 29651 CTGCTGCATG CtTTCGTCTGA TCGTGGCGCG CAACGTCCG GATATCTGAT  
 29701 CGGgAGAGAC ACCGGTGTCTG TGACAGACCT GTCGGAAGCC GaCGGCGTCG  
 29751 GCGCGGGCCG ACCAGGACAG GAAGGTAAGG AAATCCACGC CCTGCAGGGC  
 29801 GTGGCGTgAG GTTCCGCCA TGTCGACCGC TCGCGTGACG ACGCGCGCCG  
 29851 ACAGCGATGT CGTCGGGACC GCGTGCAGAT CGAGCTCCTT GGCGACGGCG  
 29901 CGGGCGATT CCGTGCCGTA GGCCAGCGCC AGCGCGCTGG AAAACATGAC  
 29951 AAGCGTGTCC TCGTCGGTGG CGACCCAGGA TACGTTGCGC CCTGACGGCG  
 30001 TCGTGCCGGA TGCGATGACC TGGAACTGCT CGCCTGCTTT GGTGACATAT  
 30051 AGGGTCTCTC CTTGGCTGGC GCGCGCGCA AAGACATCAA GCTCCAAAGC  
 30101 AGGTAAAGCG GGTGGGATCT GGAAGTTCAT GCGGTGGGCC GTTGTCTCGA  
 30151 ATCCTTGGGT GTATCGCGTT CTACGGAGCG GGAACATGGA ATATGCACTG  
 30201 TGTGACGCTT TCGGCTCTTC GGTTCATGC CGGCATGACA AACCCGACGG  
 30251 CATGCAGGCC TCGTCCCGT CCGGCCACCG AGGCTCGCCT TGAACGCTGC  
 30301 GCAAAGTGTC GGCCATACGC TGGGCCTCGA GCTGGAAATG GTGGTGGCcT  
 30351 GTCGTGCGAC CGGGGCCAGC CATCCGGTGG CGCGCCATTT CGAAGCGCTG  
 30401 CGCAACTTGC GCCGTCAACG CGGAGAGTCC GTGCAGGAGT ACCGTCTGGA  
 30451 CGCGCGCCTG TGCGGGGGGT GGGCGGTCCC CATGGCCTGA GCGGGCTTGA

Fig. 5 20/23

**THIS PAGE BLANK (USPTO)**



30501 TAACGGCTAT AACCTGCTCG AAACCGCATT CGCGCCTGTG AACGGCGGGCG  
 30551 CGGGAGGGCT GCGCCGACTG GCCGAAGCGG TGCGCCGCGA GCTGCGTGAC  
 30601 ACGCAGCTCG CCCTGGCCGC GGAGGGGGCG ATGCTGATCA ACGCGGCCGA  
 30651 GCACCCCGCC GCCAGCCTGG ATGCCGACTG GTACCGGCGA GTGCGGGTGC  
 30701 CGCGGCCCAT CTACGAGGAA CTCGTGCGCC AGCGAGGCTG GCTGCACCGG  
 30751 ATCGGGATAG ACGCCAAGGC ACAGAACAGC CCCTGCACGT CGGTTCCCGT  
 30801 GGCCATCGCC GCGCGCTGCC TGAACGTCGT GCTGGCGCTG GCTCCCGCGC  
 30851 AGATCGCCAT GTTCGCCAAC AGCCCGCTGG AGGCAGGGCG GGTGACCGGT  
 30901 CTCAAGGAAA ACCGCCTGAC CCTGTGGCCG CGCATGTTCC GAGGCGCGCG  
 30951 CTACCTGGGC GACGACCTGC TGCATCGCCT GCCTGCAAGG CCGTTTCGCG  
 31001 ATCTCGGCGA TTATTTCCGC TGGATGTTTCG GCGGATTGAC CGCCAGCCGG  
 31051 GCGCTACCGC CGGGCGACGC TTGCGACTAC AAGAACGCCG ATGTGGCCTG  
 31101 CCTGGTGGGA GCCCCTTCGC TGGCAGAGTT CCTGTATGCG GGCGCGTGGT  
 31151 CCGCGCGAAA CCTGAATGAT GGCGGTTCGG TGCCTCTGGC CGCGCGCAGC  
 31201 GAACATTTTCG TCTATTCGCA GTTCGCGCAG TTCCTGGACG CGCGTTGGCG  
 31251 CTACAGGATG CCGATTGTCC CCGCCTTGCC GGCGCTGTTG CGAGCCTGGG  
 31301 ACAGGCAGGG CGGCCTGGAA GCGCTGTTTCG AGCAGGCCGG CGCGCAAGGC  
 31351 TACATCGAGG GGCGCGCGCC GGGCGCGGTA TTTGCCGATG CCGACTTGCT  
 31401 GAGCTCAGCC GGCGATGCAG TCGCGGCCAG TGCGCCGATG GCGGCGTCGG  
 31451 CGCTGCAATT GGGGCTGTTG CGCAATCTGC ACGACGCCGA GGCCCTGGTG  
 31501 AGGCGATGGG GCTGGCTGCG CTTGCGTGCG TTGCGCGATC GGGCCATCGC  
 31551 TTTGGCGTTG GACGATGCGC AGGTGCGCTG CCTTTGCCAA CAGGTCGTGG  
 31601 CGGTAGCCGA AGGCGGGCTG GCCGGCGACG AGCAGCAATG GCTCGATTAT  
 31651 GTGCGTTACG TGGTGAAAC CGGCGAGACC GCCGCGGACC GCATGCTGCG  
 31701 CTTGTGGCGC CAGGCGCGCG GCACGCCTGA GATGCGCCGC GCACAGGCGT  
 31751 GCCGGCAGCG CGCGGTGCTG TCCTAGGATC CGGCACATTC CTTGCCAGGT  
 31801 CTCGTTGGCC AGGCCTGTCA GTGCTCCGCT TCGTACACCT CGTCGGCCGC  
 31851 CAGCAGAATA TCCACCGGTG GGTCGAAGCC GATGCGCCGC GCCGTTTCCA  
 31901 GATTGATGGC TATCTTGCGG GGGGCATTCC AGACCTGGCT GATGCTGCGC  
 31951 GGCTTTTCCC CATTGAAAAT GCGGGCAATG GTCTGGGCGT GGAACATGCC

Fig. 5 21/23

**THIS PAGE BLANK (USPTO)**

32001 TACGCTGGAG TAGTCCGCCT TGGCCAGGCT CATCAACAGG CCGGCCTTGA  
32051 CCTCGTCGGA GCCTTGATC GAGAACTCG GCACGCGGC GCGCGCAGC  
32101 AGCGCGGCGA GCTGCTTGAC GGACGTCGAG GTGATGCCCC GGTGCTCGGT  
32151 GACGTaAAAG GCGTCGACTT CGcTCGACAG CTTCTGGTAG CAAGCCAGAA  
32201 CGTTCTGGGT TGCCGTGGCG ATGGGGATGC CGGTCGCGCG TCGTCGcAA  
32251 CGCTTGACGG AGAAATCCAA TGCCGGCATT AGTGCGGCGA CCTTATCGAT  
32301 GGCTGCGTAG GTGCGACCTG CTTCCGTGTC TTCGTAGACC AGTCCAAGCG  
32351 TCTTGAAcCGG CACGATGTCA TGGAGCAGCT GGATCTGCCG CTGGTAGTGG  
32401 TCGGgCTGTA CCCGGGCATG CAGGTTGTCC TGGCCGCTGT CGGCCGCACT  
32451 GGGTATGATC CGGGCGCTTA TcGGGTCGgT CGACGAGACG ACCACGGTGG  
32501 GTACCGGCGT GCCAGTTCG ACCATGTCCT GTCCAGCCCCA GGTACCCATG  
32551 GCGATGATCA GGTGATGTC CTTGGCGCCA TGCAGGCGTg CCGCAACGGC  
32601 TTCGCGCACG GCAGGCCGCA AGGcGGTGTC GAAGTTGCCG GGCTgCCACC  
32651 ACGCATCGGG CaCGAACTCG ATGTAGTTGC TCGGGCATG CGTGCCAGG  
32701 TAAAGCCAGG CCTTTCGCAT ATCGGTTATC TCGGGCATGT CGTCGATACG  
32751 CAGCCATCCG AGTTGTTGCA ATGCGCGCGC GATCGCGTAG AGCGTGCGCG  
32801 GATACTCCTC GTACTCGCCG CTACCCACAT AACCAGTGCG CCATTTCCGG  
32851 CCGGATGTAT GGGAGGGAGG CGGAAGGCGG GGCGAGGTAA GGGCGACGCC  
32901 GGAGCTCAGG GCCGCGACAG GAGGAGGGCT GGATGCCGCC GCGATGGGCC  
32951 ACGCGAGGCC AAGAAGCAGG GCGAGCGGGG CGAGTATTCC GGGGCGAAGG  
33001 GTCATGGGCG ATGAATGGCG ATGATGGTGA GATCGTCGGA TTGTTCCGAA  
33051 TCGGCGGCGA ATTCGCGTAG GTCGTGCAGA ATGTGCTCGA TGAGTTCCGG  
33101 CGCTGCGTGC GCGCGGCCCT GCATCAGGGC GACCAGCCGC GGCAGACCAT  
33151 ACTGGGCGCA GCCGCCGTGG ATGGCTTCGG TGACGCCGTC GGTAAACGCG  
33201 ACCAGCGAGG TGCCGTTCGG CAAGGTGGTG CTCAGGGTGG AATACGCCTC  
33251 GTTGTCCAGC ACGCCGAGG CCGCGCCGCT GCTTCCTTGA AGCAGGCGGA  
33301 CCTCGCCACG TTCGTGATG AGCAGCGGCG GCGGGTGGCC GCGTTGACC  
33351 CAgGCCAGGG CGCCTGTTT CCGGGTGAAG ACGCCTATCA GCAAGGTGAC  
33401 AAACATCAGC TTGGGGTTGT TCTCGGCCAG ACGGTGGTTC ACCTGGTGG  
33451 CGATGGCGCC CGGGTCGTGC TCTTCTCCG CCACGCTGCG TATCAAGGTC  
33501 CTGACGATGG CCATGAACAG GGCCGCGGGC ACGCCTTTTC CGGATACGTC

Fig. 5 22/23

**THIS PAGE BLANK (USPTO)**

33551 GCCGATGGCA AAGCACAGAC GCCCGTCTGC CAGCACGAAG TAGTCGTAGA  
33601 AATCCCCACC GACCTCCCGG GCCGGGTACA TGACGGCACG CAACTGGCTG  
33651 CCGCGCGTGG CCGCATCGGG CAACGGCTGG GGAAGCAGGC CAAGTTGGAT  
33701 GGAGCGGGCG ATGCTCAATT CGCTTTCGAG GCGTTCGCGG TTCGATATCT  
33751 GCGCCATCAG GGCCCGCACA TTGTGGTGCA GCTGTTCGTT CATGAACAGG  
33801 AACGATTCGG CGAGCTGTCC GACTTCGTCTG CGCCGCCGGC GCGGCAGGCA  
33851 TGCCACCGAC GCGGGAACCC GGATCGGCTC GGTGAGGTCC TGGGTGGGAA  
33901 GCTGGCGAGC GTAGTTGCTC AGTTGCGCCA ACGGCCGGGC GATGCGCACC  
33951 GCCACCACCC ATGCCAGCAT CAGCCCGGCC AGCAAGGTGG CGGCGAAGAT  
34001 CAGTGCCTGC CGGCGCACCA GATTCTGTGC CGGGTCGGTC AGGTCCGGCT  
34051 CGGGAACGAC ACCGATGATG GTCCAATGCA GCGGCTTGTA TCGCAGGGCG  
34101 TCGATCTGCC AGGCGCTTTC GCCGTTGGTA AAGCGCAACG TCAGGCCGCG  
34151 GGTAGACGAG ATTTTCGGCAA GCATCGAATG CAATACCCGT CCCGATTCTGA  
34201 CGTCTGTCTGA GTCCAGCAGC CGGGCGGCCG ATGGGGGTGG CGGCACGATC  
34251 ACCGTGCCAT CGTCCGCAAC CACGAACACG AAACCATGGC GGCTGAGCCG  
34301 CAGCTCCGAC AGGTTCCGGT CTATCGCGGC AATCATGTTG GCTTTCTGGG  
34351 CGGCAACCTT GTCGATGATG GCTTGTGAGC TATCGGAGAT GGCGAGAACC  
34401 CACTTCCACG CGGGGAAGTA CACGAAATAG GCGTGTCTGCA TCTGGGCGGA  
34451 CTCGTCCAGG GGAGACGGGT AGATGGCGAA GCCGCGACCG TCGTTGCGGC  
34501 TTTCTCTGTA CATGGCGGCG GCGAGCGGCC GGCCCTTGAA GTCGCGGATC  
34551 CCgagAGGt CCCGGTCGAn nnnnCGGGG TTGGTGCTGG CCAGCACnnn  
34601 nCCTTcCGCG TCATAGGCGA AGGCGACGCG GCGCGGTCCC AGGTCGAGAT  
34651 GGTTACAGCA GACACGCGCC ATGCCCTTGG CGGCGCCGGT AGTGACGTGT  
34701 CCGCGCTCGG CCTGCGCGGC ATAGGCGTTC AGCACCGATG TCACGACCGC  
34751 GCTGAGTTGG ATCAGTTGCC TCGGGCTTTC CCGGATGGTG CGGATCTTGT  
34801 CGTCGAGCAG CGTGACCAG CGCGTGTCGG TGTCACGAAC TACCAAGTCC  
34851 ATGATGTTAC TGACGGCATG CAGTTCGTTC TTGATGATGT TGTTCGTGAC  
34901 ATCGCGCTGG GTCACGAGCA TCACGACGAT GCCTACGAGT AGCAGCGTGG  
34951 ATGCAATGAG CAGGAGGAAC TTCCCACGCA ATGAAAGCGG CAACTCTAGC  
35001 CGGGGAGTAC GGCGCATGAA CATGAA

**THIS PAGE BLANK (USPTO)**